Comparative studies of the nucleotide sequences of metallothionein-like plant genes

Bartley, Julia

Comparative studies of the Nucleotide Sequences of Metallothionein-like Plant Genes

by

Julia Bartley (née Bryden)

thesis submitted for the degree of Master of Science
to the University of Durham

Department of Biological Sciences

June 1994

ABSTRACT

Metallothioneins (MTs) are low molecular weight, cysteine-rich, metal binding proteins, whose proposed functions include roles in essential trace metal homeostasis and the detoxification of certain non-essential metals. Genes encoding MTs have been isolated from a wide range of organisms and "MT-like" genes have also been identified from several plant species. At present the function(s) of these plant genes is/are unknown. It appears that plants possess a complement of related genes which may perform different functions, possibly in different organs, and hence appear to show different patterns of expression. Unlike the E_c protein from wheatgerm, the translational products of other plant MT-like genes remain to be purified from plant tissue. Previous to this study, there were only two reports which presented 5' flanking sequences of MT-like genes from plants (maize and pea). The aim of this research was to determine the sequences of the 5' flanking regions of several different MT-like genes from several plants and to identify any consensus sequences which could be regulatory elements that control the expression of these genes.

PCR was used to amplify the coding regions of two additional MT-like genes from pea designated $PsMT_B$ and $PsMT_C$. A partial cDNA from *N. tabacum* , previously isolated via heterologous probing with $PsMT_A$, was used to screen a genomic library. Three genomic clones were identified (clones 1, 2, and 13) but subsequent sequencing failed to identify any cysteine-rich motifs. However, clone 13 showed significant sequence similarity to a gene from *A. thaliana* (*meri 5*) which is known to be expressed in apical meristems. Two MT-like genes, designated *AtMT-1* and *AtMT-2* with consideration to the location of the encoded cysteine residues, were isolated by screening an *A. thaliana* genomic library with cDNA probes generated by PCR based upon known cDNA sequences (*MT-1* and *MT-2*). The 5' flanking sequences of both genes were characterised, *AtMT-1* via sub-cloning and "gene walking" and *AtMT-2* via direct sequencing of cosmid DNA. A PCR product was also amplified using primers designed to a third MT-like cDNA from *A. thaliana* (which shows some sequence similarity to E_c from wheat) although corresponding genomic clones remain to be isolated.

The 5' flanking sequences of MT-like genes from maize, pea and a class I MT from *N.crassa* were compared with the novel 5' flanking sequences obtained for the 2 new genes from *A. thaliana* . The motif TCGCCA$(N)_{2-4}$ AATTTG has been identified, at a similar distance from the putative TATA box within the 5' flanking sequences of the MT-like genes from pea and *A. thaliana* ($PsMT_A$ and *AtMT-1*). This motif has been identified in two plant species which have been catergorised as type 1 MT-like genes, are both from dicotyledenous species and are thought to be expressed in roots in response to copper ions. Several other (short) conserved motifs have also been identified following a series of comparisons of the first 300 nucleotides of 5' flanking sequence adjacent to the putative TATA boxes of these genes.

In the course of these studies automated DNA sequence analyses have been optimised with respect to DNA template preparation and the direct sequencing of PCR products and cosmid DNA.

MEMORANDUM

Part of the work presented in this thesis has been presented in the following publication

Robinson, N.J., Evans, I.M., Mulcrone, J., Bryden, J., Tommey, A.M. (1992) Genes with similarity to metallothionein genes and copper, zinc ligands in *Pisum sativum* L. Plant and Soil. **146**: 291-298.

STATEMENT

No part of this thesis has been previously submitted for a degree in this or any other university. I declare that, unless otherwise indicated, the work presented herein is entirely my own.

ACKNOWLEDGEMENTS

I would like to thank Professors D. Boulter and P.R. Evans for allowing the use of the facilitiies in the Department of Biological Sciences. I would also like to thank my supervisors Drs J.A. Gatehouse and N.R. Robinson for their confidence, patience, expertise and guidance thoughout this study. I am grateful to Mr J.R. Wilson for the PCR and cloning of *MT-2* cDNA from *A. thaliana* and database searches, Dr D.P. Bown and A.M. Tommey for providing cDNA and genomic libraries and for their expert advice and assistance, Mr L.N. Gatehouse for advice on sequence data analysis, Mr J.S. Gilroy for oligonucleotide synthesis and Mr. D. Hutchinson for photographic work.

My sincere and heartfelt thanks go to members, both past and present, of the Biological Sciences department for their continued support, encouragement (especially in the final stages) and for not 'hounding' me for DNA sequence data! A special thank you to Miss V.E. Parker for helping to run the DNA sequencing service.

Special thanks to Paul, my family and friends for their moral support and encouragement (especially in recent months) it has been greatly appreciated.

ABBREVIATIONS.

amp	ampicillin;
bp	base pairs(s);
BSA	bovine serum albumin;
carb	carbenicillin;
cDNA	copy DNA;
cfu	colony forming units;
C-terminal	carboxy-terminal;
Cys	cysteine;
DMSO	dimethyl sulfoxide;
dATP	deoxyadenosine-5'-triphosphate;
dCTP	deoxycytidine-5'-triphosphate;
dGTP	deoxyguanosine-5'-triphosphate;
dTTP	deoxythymidine-5'-triphosphate;
DNA	deoxyribonucleic acid;
DNase	deoxyribonuclease;
EDTA	ethylenediaminetetra-acetic acid (disodium salt);
FF	Feltham First;
IAA	iso amyl alcohol;
IPTG	isopropyl-β-thiogalactopyranoside;
LB	Luria-Bertani;
MRE	metal-responsive element;
MT(s)	metallothionein(s);
N-terminal	amino-terminal;
OD_{260}	optical density at 260nm;
ORF	open reading frame;
PEG	polyethylene glycol;
PCR	polymerase chain reaction;
pfu	plaque forming units;
PPi	tetra-Sodium pyrophosphate;
RNA	ribonucleic acid;
RNase	ribonuclease;
rpm	revolutions per minute;
SDS	sodium dodecyl sulphate;
SSC	saline sodium citrate;
TE	Tris-EDTA;
Tet	tetracycline;

Tris	tris(hydroxymethyl)methylamine;
uv	ultra violet light;
w/v	weight for volume
Xaa	an amino acid other than cysteine;
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside;

CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Metallothionein

Metals, such as zinc and copper, are essential components for a variety of processes in cellular metabolism but can be toxic when present in higher concentrations. Other metals, such as cadmium and mercury have no known biological functions and are toxic even at relatively low concentration. Therefore, it is important for all organisms to balance the cellular concentration of these potentially toxic metals. Metallothioneins (MTs), are low molecular weight, cysteine-rich, metal binding proteins, which usually lack aromatic and histidine residues. Cysteine residues are typically present in the protein as Cys-Cys and Cys-Xaa-Cys motifs (where Xaa is an amino acid other than cysteine) and which bind the metal ions via thiolate bonds. The proposed functions of MTs include roles in essential trace metal homeostasis and the detoxification of certain non-essential metals (Kägi and Schäffer, 1988), but other roles have also been proposed (eg. scavenging of free radicals, modulating activity of zinc requiring transcription factors).

The first metallothionein, a cadmium-binding protein, was isolated from equine kidney cortex in 1957 (Margoshes and Vallee, 1957). MTs have since been isolated from a wide range of vertebrates, invertebrates, fungi and cyanobacteria (Karin and Richards, 1982; Maroni *et al.*, 1986; Münger *et al.*, 1987; Robinson *et al.*, 1990).

Classification and Structure of MTs.

On a structural basis it has become necessary to sub-divide metallothioneins into 3 classes (Fowler *et al.*, 1987).

class I: Polypeptides with locations of cysteine closely related to those in equine renal MT;

class II: Polypeptides with locations of cysteine only distantly related to those in equine renal MT.

class III: Atypical non-translationally synthesised metal-thiolate polypeptides. This includes secondary metabolites from higher plants, eukaryotic algae and certain fungi.

Class I and Class II MT.

Class I and II MTs are proteins which are encoded by structural genes and have been identified throughout the animal kingdom (Kägi and Kojima, 1987) in some fungi (*Neurospora crassa* , Lerch and Beltramini, 1983; *Agaricus bisporus* , Münger and Lerch, 1985) and in cyanobacteria (Olafson, 1988).

Class I MTs, in particular mammalian forms, are characterised by a molecular weight of 6000-7000, with some 61-62 amino acid residues, containing 20 cysteine (Cys) residues, 6-8 lysine residues, 7-10 serine residues, a single acetylated methionine at the N-terminus and no aromatic or histidine residues.

The linear sequence of cysteine residues are the same for class I MTs and (by definition) are closely related to that in equine renal MT. All Cys occur in the reduced form and the metal ions are bound in clusters of four and three ions associated with cysteine thiolate ligands. The structure has two domains, the N-terminal domain which binds the three metal ion cluster, and the C-terminal domain which binds the remaining four metal ion cluster (Furey *et al* ., 1987).

Class I MTs display extensive genetic polymorphism (Kägi and Kojima, 1987). Mammalian tissues usually contain two or more distinct isoforms grouped into two major fractions, MT-1 and MT-2, differing by a single negative charge at neutral pH. In the mouse, the genes encoding these two isoforms are ca. 6 kb apart on chromosome 8

(Searle *et al.*, 1984). In many cases, there are subforms within these fractions, separable by HPLC and specified as MT-1a, MT-1b etc. and MT-2a, MT-2b etc. (Fowler *et al.*, 1987). The various isoforms exhibit small differences in their binding affinity for metal ion ligands, but otherwise appear to have similar biochemical properties. The multiplicity of MTs has been confirmed by gene cloning experiments showing that MTs are encoded by a multigene family (Hamer, 1986). Two other members of the MT gene family (at least in humans and mice) have been identified and designated MT-3 and MT-4. MT-3 was cloned following the characterisation of its product in humans, a protein that was shown previously to inhibit neuronal survival in culture and to be deficient in the brains of people with Alzheimer disease (Palmiter *et al.*, 1992).

Different isoforms of human MT have shown differential transcription responses: The hMT-1_A promoter is responsive only to cadmium, whereas the hMT-2_A promoter is responsive to zinc and glucocorticoids as well as cadmium (Richards, 1984). Unlike MT-1 and MT-2, which are expressed in most organs, MT-3 expression appears to be restricted to the brain (Palmiter *et al.*, 1992) and MT-4 is only expressed in tongue (unpublished data cited in Masters *et al.*, 1994).

In animals the protein is most abundant in parenchymatous tissues, ie. liver kidney, pancreas and intestines. There are wide variations in concentration in different species and tissues, reflecting age, stage of development, diet and other not yet fully identified factors. Although MT is generally a cytoplasmic protein, it can also accumulate in lysosomes usually as a copper -MT (Johnson *et al.*, 1981) and it has also been observed in the nucleus during development (Nartey *et al.*, 1987) and in the nucleus of cultured rat cells at certain stages of the cell cycle (Tsujikawa *et al.*, 1991).

Class II MTs contain similar proportions of metal-binding cysteine residues, mostly in Cys-X-Cys motifs, but the locations of the cysteine residues are only distantly related to those in equine renal MT. Class II MTs have been identified from a number of organisms, eg. sea urchin (Nemer *et al*., 1985), yeast (Premakuman *et al*., 1975) and cyanobacteria (Olafson *et al*., 1988). A zinc-binding protein (E_c) has been isolated from wheatgerm (Lane *et al.,* 1987), and designated class II MT (Kägi and Schäffer, 1988), due to its abundant Cys-X-Cys motifs and metal-binding ability. Equivalent proteins have not yet been purified from vegetative plant tissues.

Class III MT.

Class III MTs are a unique family of metal-binding polypeptides, which form aggregates in the presence of metals. They were discovered by workers searching for evidence that plants and fungi contain proteins like mammalian MTs. However, such metal-binding complexes are aggregates of a heterogeneous population of polypeptides of differing lengths and are known by a variety of names including: cadystin (Murasugi *et al*., 1981); phytochelatin (Grill *et al*., 1985); gamma-glutamyl metal-binding peptide (Reese *et al*., 1988); phytometallothionein (Rauser, 1987) and poly(γ-glutamyl cysteinyl)glycine or $(\gamma\text{-EC})_nG$ (Robinson and Jackson, 1986; Jackson *et al*., 1987). The term phytochelatins is most commonly used, but the 'phyto' part may become redundant following recent reports of equivalent metal-binding polypeptides from animals (N.J. Robinson, personal communication).

The first class III MT to be identified was in extracts from the fission yeast *Schizosaccharomyces pombe* (Murasugi *et al.*, 1981) where they were found to be inducible by cadmium and termed cadystins. Similar polypeptides were subsequently purified from plant cell cultures (Grill *et al*., 1985) and termed phytochelatins. The ability to synthesise phytochelatins in response to toxic trace metals is conserved from the most advanced group of higher plants, the Orchidales, to the more primitive red, green and brown algae (Steffens, 1990).

These polypeptides are products of secondary metabolism, and have the primary structure of repeating gamma glutamylcysteinyl units, $(\gamma\,\text{Glu-Cys})_n\text{Gly}$, where n=2-11.

The peptide bonds between Glu and Cys in the repeating Glu-Cys pairs are of the g-carboxamide type. The presence of these gamma-carboxamide bonds was confirmed by analysis of ^{13}C NMR spectra obtained using isolated polypeptides labeled with L-(1,5-$^{13}C_2$) glutamic acid (Jackson *et al*., 1987).

In comparison to phytochelatins, glutathione (GSH) possesses the structure γ Glu-Cys-Gly, in which the peptide bond is formed between the gamma- or side-chain carboxylate of glutamic acid rather than alpha-carboxylate utilised in peptide bonds of polypeptides whose synthesis is ribosome-dependent. This indicates that their biosyntheses share a common enzymology. Buthionine sulfoximine (BSO), a potent inhibitor of the enzyme gamma-glutamylcysteine synthetase, the first enzyme of the GSH biosynthesis (Glu+Cys > γ Glu-Cys) also inhibits phytochelatin synthesis. This suggests that GSH, or its precursor gamma Glu-Cys, act as direct phytochelatin precursors. Additional evidence for the participation of GSH in phytochelatin biosynthesis derives from the identification of (γ-Glu-Cys)$_n$-β-Ala in the *fabales* (Grill *et al*., 1987). These species accumulate homo-glutothione in which glycine is substituted with β- alanine, rather than glutathione and in response to metals accumulate an equivalent analogue of phytochelatin (termed homophytochelatin). Early reports of the *in vitro* analysis of the enzyme gamma-glutamlycysteinyl dipeptidyl transpeptidase (Grill *et al*., 1989), the putative phytochelatin synthetase which assembles the polymeric aggregates of gamma glutamylcysteine, have not been succeeded by detailed characterisation of the enzyme. Thus, although it is clear that these peptides are the product of a biosynthetic pathway, the final stages of this pathway remains to be fully elucidated.

MT gene organisation and expression.

Gene expression is, at least in part, controlled by DNA-binding proteins (*trans* -activating factors) which recognise specific sequences in the 5' flanking regions of genes (*cis* -acting control sequences), and interact with a complex of RNA polymerase and various transcription factors required for active gene transcription. MT gene expression is rapidly and transiently induced by exposure to trace metals. In mammalian cells, various circulating factors such as hormones and interferon are also known to act as inducers. Mammalian MT genes are also transcriptionally activated by a variety of environmental stimuli other than trace metals (Palmiter, 1987). In all cases the rate of protein synthesis was proportional to the amount of MT mRNA, suggesting transcriptional control, although small changes in MT mRNA stability have also been suggested in some studies (Mayo and Palmiter, 1981).

Mammalian MT

The optimal metal ion concentration for induction varies in different systems but is generally just below the level of toxicity. The upstream MT gene regulatory sequences involved in trace metal induction have been studied in experiments in which mutated or hybrid control sequences are constructed by recombinant DNA techniques and reintroduced into cells. The *cis* -acting control sequences have been studied for the mouse MT-1 (Mayo *et al*., 1982) and human MT-II_A genes (Karin and Richards, 1982). The two systems were found to be regulated by similar mechanisms (Hamer, 1986). Analysis of 5' deletion mutants revealed that 60 bp of DNA was necessary for metal induction, although 3' deletion mutants, which were lacking the 5' flanking region up to postion -126, position 1 being the start of the initiation codon, showed some metal regulation. It was found that regions outside these metal regulatory regions could also play an important role in determining the efficiency of transcription. There were homologous sequences which determined a basal level of transcription in both mouse MT and human MT. It has been suggested that these regions act as enhancer elements (Hamer and Khoury, 1983). The glucocorticoid and interferon regulatory elements of these genes have also been identified. It is notable that the metal regulatory elements of MT genes in a variety of organisms, like other regulatory elements of many other eukaryotic genes, are present in multiple copies (Searle *et al* ., 1985; Hamer and Khoury, 1983).

The search for cellular regulatory factors has proven to be less "straight forward" than the identification of the DNA sequences with which they interact. Metal regulatory elements (MREs) have been identified in the promoter regions of a number of different MT-genes (Carter *et al* ., 1984; Karin *et al* ., 1984; Stuart *et al* ., 1984; Anderson *et al* ., 1986; Karin *et al* ., 1987; Otto, 1987; Harlow *et al* ., 1989; Zafarullah, 1988). There are six non-identical copies (MREa to MREf) present within the first 200 bp 5' of the transcriptional start site of the mouse MT-1 gene (Carter *et al* ., 1984; Karin *et al* ., 1984; Stuart *et al* ., 1984; Stuart *et al* ., 1985; Mueller *et al* ., 1988). MREa-d confer metal-responsive transcription when tested independently in front of a reporter gene. MREd was shown to be the strongest element for metal induction (Stuart *et al* ., 1988; Searle, 1990) and has a capacity to respond to the same spectrum of metal ions (cadmium, zinc and copper) as does the complete gene promoter, suggesting that all MREs are responsive to the different metals and together act to facilitate a strong induction response (Culotta and Hamer, 1989).

Competition experiments suggested that one or more positively acting transcription factor(s) interact with these elements (Seguin *et al* ., 1984; Scholer *et al* ., 1986). A mouse nuclear factor of 108 kDa, designated MEP-1, binds with high affinity to the MREd of the MT-1 gene (Seguin and Prévost, 1988; Labbé *et al* ., 1991). To determine if different MREs can bind common nuclear factors, competition experiments in an exonuclease III footprinting assay were performed. MEP-1 was purified to homogeneity from the metal resistant mouse L50 cells and its binding properties characterised (Labbé *et al* ., 1993). Footprinting studies demonstrated that purified MEP-1 specifically binds to MRE sequences and that it is sufficient to produce a specific footprint on the mouse MREd (Labbé *et al* ., 1991). To assess the binding properties of MEP-1 to different MREs, protein blotting and UV cross-linking experiments were used. It was observed that MEP-1 not only bound with high affinity to MREd but also to other MRE sequences present in the promoter of the mouse MT-1 gene with affinities that are proportional to their relative transcriptional strength *in vivo* . MEP-1 also binds to MREs of the human MT-II_A and trout MT-B genes (Labbé *et al* ., 1991). It was also observed that DNA-binding activity of the protein(s) interacting with the MREd is/are inactivated by the chelating agent 1,10-phenanthroline *in vitro* and can be restored by zinc ions, to support the idea that the MREd-binding protein(s) is/are important for metal-inducible transcription (Séguin, 1991). No other cations tested restored activity to the chelated protein. Therefore, MEP-1 requires zinc ions for its specific interaction with MREd and it was demonstrated that a single nucleotide substitution introduced in the core MRE region at a nucleotide required for *in vivo* transcriptional activity of MREd (Cullota and Hamer, 1989) completely abolished MEP-1 binding activity as assayed by competition experiments.

An MRE-binding factor, designated MTF-1 (MRE-binding transcription factor) which binds to the mouse MREd sequence was demonstrated in HeLa cell nuclear extracts to be inactivated by removal of zinc ions and reactivated by addition of zinc ions (Westin and Schaffner, 1988). A mouse cDNA encoding the MREd-binding protein MTF-1 (presumably synonymous with MEP-1) has subsequently been isolated and characterised (Radtke *et al* ., 1993). MREd was used to screen a cDNA expression library for MTF-1. However, MREd also binds the ubiquitous transcription factor Sp1. Therefore, using a compilation of known MRE sequences an MRE oligonucleotide (MRE-s) was designed which had the same high binding affinity for MTF-1 but lacked any Sp1-binding activity (Radtke *et al* ., 1993). A mouse lymphocytic leukemia cell line Lambda gt11 expression library was screened with an oligonucleotide probe containing multiple copies of MRE-s. One clone was found in 2.0 x 10^6 plaques, characterised and designated mMTF-1a. This clone encoded a protein of 69 kDa containing six zinc fingers of the TFIIIA type (C_2H_2) (Brown *et al* ., 1985; Miller *et al* ., 1985), followed by a putative activation domain with a high representation of acidic amino acids, followed by proline-rich sequences (Radtke *et al* ., 1993). Another MTF-1 cDNA was also isolated from a library derived from mouse lung epithelium and encoded an MTF protein with an extra 25 amino acids at the N-terminus. Many mammalian transcription factors are encoded by families of related genes, or are varied by alternative splicing. It is yet to be established whether there are multiple MTF-1 genes (Radtke *et al* ., 1993).

Fungal MT.

Similar deletion studies to the ones mentioned above were employed to identify the *cis* -acting control sequences of the *CUP* 1 gene (Butt and Ecker, 1987), which encodes MT on chromosome VIII, in the yeast *Saccharomyces cerevisiae* . Transcription of the *CUP* 1 gene increases following exposure to elevated concentrations of copper and silver ions in the growth medium (Karin *et al* ., 1984; Butt *et al* ., 1984). Transcriptional induction is mediated through the action of a *trans* -activating factor, designated ACE1/CUP2, encoded on chromosome VII (Thiele and Hamer, 1986; Welch *et al* ., 1989; Buchman *et al* ., 1989; Casas-Finit *et al* ., 1992; Thiele, 1992). In the apoprotein form, ACE 1 cannot bind to DNA. In the presence of the copper or silver ions, the N-terminal domain undergoes a conformational switch into a folded, protease-resistant form that specifically recognises the *CUP* 1 upstream *cis* -acting control sequence (Culotta *et al* ., 1989). Spectroscopic studies have shown that the N - terminal region directly binds copper in a "Copper- fist" conformation.

Prokaryotic MT

Proteins termed 'MT-like' have only been described in two prokaryotic groups, *Synechococcus* sp.(Maclean *et al* ., 1972; Olafson *et al* ., 1980; Olafson, 1984; Olafson, 1991; Takatera and Watanabe, 1992) and *Pseudomonas putida* (Higman *et al* ., 1984). Class II MTs have since been isolated from a number of *Synechococcus* strains of cyanobacteria and the MT from *Synechococcus* TX-20 was subsequently sequenced (Olafson *et al*., 1988).

DNA from *Synechococcus* PCC 6301 was used in a standard polymerase chain reaction and in a ligation-mediated polymerase chain reaction to generate products corresponding to part of an MT gene, the products were sequenced and the gene called *smtA* (Robinson *et al*., 1990). These fragments were subsequently used as probes to isolate an MT divergon, *smt*, which includes *smt A*, and a divergently transcribed gene *smt B* (Huckle *et al*., 1993). SmtA is identical to the polypeptide previously characterised by Olafson *et al*. (1988), with the exception of a serine substitution for cysteine 32 and two additional amino acids at the C-terminus (histidine and glycine). A gene from *Synechococcus vulcanus* encoding a polypeptide with similarity to SmtA, designated *mtnA*, has also been identified within the sequences flanking a previously characterised gene, *psaC* (Shimizu *et al*., 1992).

MT in *Synechococcus* sp. was shown to increase in abundance following exposure to elevated concentrations of cadmium or zinc, but not copper and the purified protein was shown to associate with cadmium or zinc (dependent upon the growth conditions) with copper as a minor component (Olafson *et al*., 1988). Similarly, partly purified MT-like protein from cadmium-exposed *Anacystis nidulans* R2 (equivalent to *Synechococcus* PCC 7942) contained predominantly cadmium with lesser amounts of zinc and copper (Takatera and Watanabe, 1992).

SmtA has been expressed in *E.coli* as a recombinant fusion protein and the protein was shown to associate with zinc, cadmium, copper and mercury ions following purification from cells grown in metal supplemented media (Shi *et al*., 1992). Enhanced accumulation of zinc was also observed in these bacterial cells, in which production of SmtA is not metalloregulated, suggesting zinc-binding *in vivo* (Shi *et al*., 1992). The pH of half-dissociation of zinc, cadmium and copper from recombinant SmtA metallothionein was determined to be 4.10, 3.50 and 2.35

respectively (Shi *et al*., 1992). In comparison to equine renal MT, recombinant SmtA had a greater affinity for zinc but lesser affinities for cadmium and copper.

Cyanobacterial (*Synechococcus* PCC 7942) mutants with an interrupted *smt* divergon, *smt*$^-$, are sensitive to zinc and show some reduction in tolerance to cadmium (cited in Turner and Robinson, 1994). These cells retained normal tolerance to copper (cited in Turner and Robinson, 1994) and mercury (Turner, 1993) indicating independence of copper and mercury resistance from *smt* -mediated metal tolerance. The *smt* divergon can be used as a marker to select for transformants derived from *smt*$^-$ cells (Turner *et al*., 1993; cited in Turner and Robinson, 1994). Cells containing re-introduced *smt* have been successfully isolated from *smt*$^-$ cells based upon restored tolerance to zinc (Turner *et al*., 1993).

The protein coding regions of *smtA* and *smtB* are divergently transcribed and separated by a 100 bp operator-promotor region (Huckle *et al*., 1993; Morby *et al*., 1993). The deduced SmtB polypeptide shows sequence similarity to a number of bacterial proteins, some of which are known to be transcriptional regulators and /or involved in metal metabolism. A divergently transcribed ORF which has been partially sequenced is present within the sequences upstream (116 bp from the ATG) of *mtnA*, from *Synechococcus vulcanus* (Shimizu *et al*., 1992). It encodes part of a protein with similarity to SmtB and the gene has been referred to as *mtnB* (cited in Turner and Robinson, 1994).

The abundance of *smtA* transcripts were shown to increase in response to elevated concentrations of a number of metal ions (including Cd^{2+}, Zn^{2+}, Cu^{2+}, Hg^{2+}, Co^{2+} and Ni^{2+}) but not heat shock (Huckle *et al*., 1993). However, at maximum permissive concentrations, only zinc and to a lesser extent cadmium and copper ions, increased expression of a reporter-gene (*lacZ*) driven by the *smtA* operator-promotor region (Huckle *et al*., 1993). In *smt*$^-$ mutants highly elevated expression of *lacZ* (driven by the *smtA* operator-promotor) was detected, even in the absence of added metal ions (Huckle *et al*., 1993). Repression, and metal-dependent expression, of *lacZ* was restored (at least in part) in cells containing plasmid borne and/or chromosomal *smtB*. SmtB is thus a *trans* -acting repressor of expression from the *smtA* operator-promotor.

Three protein complexes with the *smt* operator-promotor have been identified by electrophoretic mobility shift assays and designated MAC1, MAC2 and MAC3 (Morby *et al*., 1993). MAC1 formed with a region of DNA immediately upstream of the ATG of *smtA* and was only observed in extracts from cells containing *smtB* (MAC2 and MAC3 were retained using extracts from *smt*⁻ mutants). It was proposed that SmtB forms the protein component of MAC1 (Morby *et al*., 1993) and this has recently been confirmed following expression, affinity purification and sequencing, of SmtB in/from *E.coli* (cited in Turner and Robinson, 1994). Helix-turn-helix DNA binding proteins generally bind to inverted repeats. A candidate SmtB-binding site is a degenerate 6-2-6 inverted repeat (TGAACA-GT-TATTCA) which also incorporates the left half of a 6-2-6 direct repeat (TATTCA-GA-TATTCA). A similar inverted repeat (TGAACA-GT-TGTTCA) is present within the operator-promotor region of the MT divergon, *mtn*, of *Synecho-coccus vulcanus*. The regions of DNA involved in the formation of MAC2 and MAC3 have also been mapped by electrophoretic mobility shift assays using specific competitor DNA fragments (Morby *et al*., 1993). The MAC2 binding site contains a 7-2-7 hyphenated inverted repeat (CTGAATC-AA-GATTCAG) while MAC3 binds to a region most distal to *smtA*.

In summary, MT genes have been found to be expressed in most tissues of most organisms studied, they are transcriptionally regulated by metals and certain hormones. Some of the DNA sequences that allow response to these stimuli have been identified in the 5' flanking region of the MT genes and transcription factors that interact with some of these sequences have been cloned and characterised from animals, fungi and cyanobacteria.

Plant genes with sequence similarity to metallothionein genes.

A new family of higher plant genes has been identified which encode proteins with some sequence similarity to metallothioneins (Evans *et al* ., 1990; de Miranda *et al* ., 1990; Kawashima *et al* ., 1991; Takahashi, 1991; Okumura *et al* ., 1991; de Framond, 1991; Klemsdal *et al* ., 1992; Robinson *et al* ., 1992; Weig and Komor, 1992; Raynal *et al* ., 1993; Zhou and Goldsbrough, 1994; Snowdon and Gardner, 1993). Computer-based searches select metallothionein as the most similar known proteins. At present the functions of these genes are unknown and unlike the E_c protein isolated from wheat germ (Lane *et al* ., 1987) the translational products of these genes remain to be purified from plant tissue.

All the genes identified have at least two cysteine-rich domains containing Cys-Xaa-Cys motifs (where Xaa is an amino acid other than Cys). These domains, which are generally conserved (in different types) in different species with respect to the cysteine residues, are separated by central 'spacer' region of up to 40 amino acids which is less conserved. Two categories of metallothionein-like proteins are proposed on the basis of the predicted locations of cysteine residues and are designated types 1 and 2 and differ within the N-terminal domain. Type 1 contain exclusively Cys-Xaa-Cys motifs whereas within type 2 there is a Cys-Cys and Cys-Xaa-Xaa-Cys pair. A possible third type, which encode the E_C protein from wheat, differs by the presence of a third domain encoding Cys-Xaa-Cys motifs situated between the N-terminal and C-terminal. A second predicted metallothionein-like protein from barley and a third predicted metallothionein-like protein from *Arabidopsis thaliana* deviate from the three patterns detailed above. The N-terminal sequence from barley encodes the same distribution of

Cys residues as type 2 but differ from all three types in the number of Cys-Xaa-Cys motifs at the C-terminal domain, whereas the N-terminal sequence of the atypical predicted product from *A . thaliana* has the same distribution of Cys residues as type 1 but there is an extra amino acid present between the second and third Cys-Xaa-Cys motifs at the C-terminal. This predicted *A. thaliana* MT-like protein also lacks the central spacer region of ca. 40 amino acids. However, due to the number and location of five out of six Cys-Xaa-Cys motifs it will be referred to hereafter as a type 1 MT-like plant gene (figure 1).

Occurence and isolation.

Early reports of MT-like proteins in plants were succeeded by reports of poly(γ-glutamylcysteinyl)glycine [(γEC)nG] (Grill *et al* ., 1985; Jackson *et al* ., 1987; Grill *et al* ., 1987), as previously discussed. It has been considered by some that all of the MT-like proteins isolated from higher plants were impure isolates of (γEC)nG (Grill *et al* ., 1987). However, following the identification of MT- like genes in plants, this presumption is questioned.

Plant genes with similarity to metallothionein genes have now been identified and isolated in both monocotyledonous (barley, maize) and dicotyledonous (*Arabidopsis* , *Mimulus* , pea, soyabean) species which suggests a broad species distribution. However, unlike class III MTs, which have been the focus of an extensive study (cited in Grill *et al* ., 1989) an extensive survey of genera has yet to be reported.

Type 1

Mimulus guttatus

A cDNA library was constructed from poly(A)+ RNA from roots of *Mimulus guttatus* exposed to copper for 24 hours. The library was then screened with first-strand

Figure 1.

Sequences of the predicted products of plant genes with similarity to metallothionein genes and of the E_c protein.

Type 1

```
             1                                                 50                 83
Maize        ...MSCSCGSSCGCGSSCKCGKKYPDL.EETSTAAQ..PTVVLGVAPEKKAAPEFVEAAAESGEAAHGCSCGSGCKCDPCNC.
Barley       ...MSCSCGSSCGCGSNCNCGKMYPDL.EEKSGATMQVTVIVLGVGSAKV....QFEEAAEFGEAAHGCSCGANCKCNPCNC.
Wheat        ...MSCNCGSGCSCGSDCKCGKMYPDLTEQGSAAAQVAAVVVLGVAPENKAG..QFEVAA..GQSGEGCSCGDNCKCNPCNC.
Pea          ..MSGCGCGSSCNCGDSCKCNKRSSGLSYSEMETTE...TVILGVGPAKIQFEGAEMSAA..SEDG.GCKCGDNCTCDPCNCK
M.guttatus   .MSSGCSCGSGCKCGDNCSC.SMYPD.....METNTTV.TMIEGVAPLKMYSEGSEKSFG..AEGGNGCKCGSNCKCDPCNC.
A.thaliana   MADSNCGCGSSCKCGDSCR..............................CEKNY..NKE.CDNCSCGSNCSCGSNCNC

Consensus    .....C.CGS.C.CG..C.C.......................GV......................GC.CG..C.C.PCNC.
Cysteine          C C   C C   C C                                                C C   C C  C C
```

Type 2

```
             1                                                 50                 83
Soyabean     MSCCGGNCGCGSSCKCGNGCGGCKMYPDLSYT.ESTTTETLVMGVAPVKAQYESAEMG.....AENDGCKCGANCTCNPCTCK
Castorbean   MSCCGGNCGCGSGCKCGNGCGGCKMYPDMSFS.EKTTTETLVLGVGAEKAHFEGGEMGVV..GAEEGGCKCGDNCTCNPCTCK
A.thaliana   MSCCGGNCGCGSGCKCGNGCGGCKMYPDLGFSGETTTTETFVLGVAPAMKNQYEASGESN..NAESDACKCGSDCKCDPCTCK

Consensus    MSCCGG.CGCGS.C.CG.GCGGC.MY.DL....E.TT..T...GV.......................CKCG..C.C.PCTCK
Cysteines      CC   C C   C C   C  C                                              C C   C C  C C
```

Ec type

```
                      1                                                 50
Wheat        .........MGCDDKCGCAVPCPGGTGCRCTSARSGAAAG.EHTTCGCGEHCGCNPCACGREGTPSGRANRRANCSCGAACNCASCGSATA
A.thaliana   ADTGKGSASASCNDRCGCPSPCPGGESCRCKMMSEASGGDQEHNTCPCGEHCGCNPCNCPKTQTQTS....AKGCTCGEGCTCATCAA...

Consensus    ..........C.D.CGC. PCPGG..CRC..........EH.TC.CGEHCGCNPC.C....T.........C.CG..C.CA.C. ...
Cysteine               C   C C   C     C C                C C   C C  C C                 C C   C C  C
```

Others

```
             1                                                 50                                      90
Barley       MSCCGGKCGCGAGCQCGTGCGGCKMFPDVEATAGAAAMVMPTASHKGSSGGFEMAGGETGGCDCATCKCGTRAAAPAAAASEPAPGRPAG
             RGEHEDERRTSNTNQAPSPSPSYHQ
```

cDNA prepared from roots of copper-treated and untreated plants. Five copper-repressible cDNA clones (J39, 49, 55, 73, 87) were identified and subsequently sequenced (de Miranda *et al* ., 1990). The clones are representatives of two distinct sequences which differ by 3 bp in the ORF (one in the termination codon) and at 13 positions outside this coding region. However, none of these changes affect the predicted amino acid sequence which encodes a protein of 72 residues. Minor differences in the length of the 5' and 3' ends demonstrate that four of the five cDNAs represent independently created clones. The five cDNA clones are divided into two classes, suggesting that there must be at least two copies of these sequences per genome.

Pisum sativum* and *Zea mays

Differential screening of cDNA libraries constructed from poly(A)+ RNA from roots of garden pea (*Pisum sativum*) and maize (*Zea mays*) (Evans *et al* ., 1990; de Framond, 1991) led to the isolation of cDNA sequences which were preferentially expressed in roots rather than other organs. The corresponding genes, $PsMT_A$ and MT-L (metallothionein-like) were subsequently isolated from pea and maize genomic libraries and the predicted amino acid sequences encode proteins of 75 residues and 76 residues respectively. Using the cDNA sequences pPR179 from pea and pCIB1325 from maize as probes, Southern analyses revealed that the probes hybridised to multiple fragments of restricted pea and maize genomic DNA respectively suggesting the presence of small multi-gene families. Partial sequences of two further members of the pea gene family, $PsMT_B$ and $PsMT_C$, have been obtained following PCR-mediated cloning and will be detailed herein (Robinson *et al* ., 1992).

Barley

A cDNA library was constructed from root poly(A)+ RNA isolated from iron deficient barley roots, the library was differentially screened with probes prepared from poly(A)+ RNA from iron-deficient and iron-sufficient roots (Okumura *et al*., 1991). A cDNA clone of about 500 bp, assumed to be full length by northern hydridisation, was identified and designated *ids- 1* (iron deficiency-specific clone 1). The predicted amino acid sequence encodes a protein of 74 residues.

Arabidopsis thaliana

A DNA sequence of a cDNA was isolated and characterised from a root cDNA library from *Arabidopsis thaliana* and entered in the GENBANK database (Zhou and Goldsbrough, accession number L15389). The predicted amino acid sequence encodes a 45 amino acid protein which is highly homologous to animal MT proteins

Wheat

A cDNA library was constructed from poly(A)+ RNA isolated from the root tips of wheat cv Warigal treated for 2 days with 10 µM aluminium (Al). The library was differentially screened with probes prepared from poly(A)+ RNA from control Warigal roots (no Al) and from Warigal roots treated for 2 days with 10 µM Al. A total of five individual cDNA clones were characterised and termed *wali1* to *wali5* (for wheat alumi-nium induced). One of the clones, *wali1*, shows sequence homology to type 1 MT - like genes from plants (Snowdon and Gardner, 1993). The predicted amino acid sequence encodes a protein of 67 residues.

Type 2.

Soyabean

Using a synthetic 21-mer oligonucleotide as a probe, that corresponded to the consensus nucleotide sequence of the N-terminal region of mammalian MT, a sequence designated 21-1-A, was isolated from a soyabean (*Glycine max*) cDNA library (Kawashima *et al*., 1991). The predicted amino acid sequence encodes a protein of 79 residues. Northern analysis of total cellular RNA isolated from roots and leaves revealed one positive band in each, which corresponded to 700 bp. The level of the transcript of the soyabean gene was found to be higher in leaves than in roots.

Arabidopsis thaliana

A DNA sequence of a cDNA was isolated and characterised from a leaf cDNA library from *Arabidopsis thaliana* using above sequences as probes and entered in the EMBL database (Takahashi, unpublished 1991; accession number X62818). The predicted amino acid sequence encodes a protein of 81 residues.

Castorbean

A sequence has also been isolated and characterised from a leaf cDNA library from castorbean and entered in the EMBL database (Weig and Komor, unpublished 1992; accession number L02306). The predicted amino acid sequence encodes a protein of 80 residues.

Type E_c

Wheat

A cDNA library was prepared from bulk mRNA of mature wheat embryos and screened using oligonucleotide probes that encoded parts of the partial amino acid sequence for the Zn-containing E_c protein isolated by Lane *et al*. (1987). Eleven cDNA clones were isolated, characterised and divided into three classes (I, II,and III), the sequences within each class being of varying length but identical in regions of overlap. The predicted amino acid sequence of all the cDNAs indicate that they encode proteins of 81 residues which are very similar to the partial polypeptide sequence previously determined for the E_c protein (i.e. 56 of the 59 amino acids previously determined for the E_c protein are identical to the sequences encoded in cDNA-I, cDNA-II and cDNA-III).

The sequence determined for cDNA-III encodes a protein that differs at three positions from the one encoded in cDNA-I and cDNA-II (positions 4, 27 and 79). The cDNA-III clone was used to screen a wheat genomic library and a 3.1 kb genomic fragment, designated gf-3.1, that encodes the same protein as cDNA-III was isolated. E_c genes isolated (Kawashima *et al* ., 1992) were found to be located in single copies on the long arms of chromosomes 1A, 1B and 1D in hexaploid wheat. However, they are unlike animal metallothionein genes which are contained in multigene clusters.

Arabidopsis thaliana

A partial cDNA sequence has been isolated from an *Arabidopsis thaliana* cDNA library of total seed and entered in the EMBL database (Raynal *et al* ., 1993; accession number Z27049) The predicted amino acid sequence showed significant similarity with respect to the locations of cysteine residues to wheat E_c (Kawashima *et al* ., 1992).

Others

Barley

A barley aleurone cDNA library prepared from poly(A)$^+$ RNA of immature grains was differentially screened for clones representing transcripts present in the aleurone but not in the starchy endosperm and a clone designated pB22E was identified. Using pB22E as a probe, corresponding clones were isolated from cDNA libraries constructed from poly(A)$^+$ RNA of aleurone and pericarp, and poly(A)$^+$ RNA of embryos of developing grains. Two positive aleurone and pericarp clones (pB22E.a12 and pB22E.a16) and two positive embryo clones (pB22E.e49 and pB22E.e11) containing inserts of 191, 541, 208 and 116 bp respectively, were identified and corresponded to the sequence of pB22E. The insert of pB22E.e49 was used as a probe to screen a cDNA library constructed from poly(A)$^+$ RNA of 2 day germinated barley. Four positive clones (pB22E.g49, pB22E.g411, pB22E.g312 and pB22E.g1) containing inserts of 397, 304, 338, and 260 bp respectively were identified. The insert from the cDNA pB22E.a16 was used as a probe to screen a genomic library and a hybridising clone was identified and designated pB22EL8 (Klemsdal *et al* ., 1991). From the cDNA sequences, two types of transcripts were identified and the predicted amino acid sequence for the two different types encode novel barley proteins of 115 and 119 amino acids respectively. The predicted amino acid sequence for clone pB22EL8 (a type 1 transcript) encodes a metallothionein-like protein of 115 residues (cited in Robinson *et al* ., 1993).

Nicotiana tabacum* and *Medicago sativum

The sequences of known MT-like genes were used as probes to identify partial cDNA clones containing homologues from other species including tobacco (*Nicotiana tabacum*) and alfalfa (*Medicago sativum*) (Robinson *et al* ., 1992) and to reveal cognates in several other higher plant species by Southern analysis.

Metal-binding characteristics of recombinant products of plant genes

The predicted MT-like proteins from plants differ from archetypal MTs in that their cysteine-rich regions (two in most cases) are shorter, but still align with particular domains of MTs, and are separated by a long intervening peptide strand (ca. 40 residues in the majority of sequences identified to date). However, in most archetypal MTs, the 2 metal-binding, cysteine-rich domains are connected by a short 'linker' region of only ca. 2 residues (Kay *et al* ., 1990; Kägi and Schäeffer, 1988).

The products of metallothionein genes from different organisms have been shown to form conformations suitable for association with metal ions when expressed in *E.coli* and yeast (Murooka and Nagaoka, 1987; Romeyer *et al* ., 1988, 1990; Jacobs *et al* ., 1989; Kille *et al* ., 1990, Silar and Wegnez, 1990). In some cases MT has been demonstrated to bind metal ions when expressed as a fusion with another protein (Romeyer *et al* ., 1988; Jacobs *et al* ., 1989; Romeyer *et al* ., 1990) However, other than the E_c protein isolated from wheatgerm which is known to associate with zinc (Lane *et al* ., 1987), attempts to isolate the translational products from MT-like genes from plants have not yet been successful. Therefore, in the absence of purified native protein, the *PsMT$_A$* gene from pea was expressed in *E.coli* to facilitate examination of the metal-binding properties of its product (Tommey *et al* ., 1991; Kille *et al* ., 1991).

The $PsMT_A$ protein was expressed in *E.coli* as a carboxyterminal extension of glutathione-S-transferase (GST) (Tommey *et al* ., 1991). A protein corresponding to the size of GST (26.5 kDa) was detected in crude lysates of induced JM101 cells containing the vector pGEX3X and a unique protein, corresponding to the predicted size (ca. 34.5 kDa) of the $PsMT_A$-GST fusion, was detected in crude lysates of cells transformed with plasmid pGPMT3. The $PsMT_A$-GST fusion protein was purified by affinity chromatography using glutathione-Sepharose 4B. A single protein of ca. 34.5 kDa was detected in fractions eluted with buffer containing 5 mM glutathione and corresponded in size to the most abundant protein observed in crude lysates of the cells containing the construct. Purified proteins from *E.coli* grown in media supplemented with cadmium, copper and zinc were used to determine the bound metal content of GST and the $PsMT_A$-GST fusion. $PsMT_A$-GST fusion bound an estimated amount of 7.07, 6.27 and 5.99 moles of cadmium, copper and zinc respectively per mole protein, compared to equivalent estimates of 0.26, 0.63 and 0.37 moles for GST alone (Tommey *et al* ., 1991). Similar estimates for Fe-binding were 0.28 moles for $PsMT_A$-GST fusion protein and 0.1 moles for GST alone (Robinson *et al* ., 1992).

Estimations of the pH at which 50 % of metal ions dissociate is a criterion used to distinguish MT from non-MT metal-binding proteins (Vasak and Armitage, 1986). The pH of half-dissociation of zinc, cadmium and copper from $PsMT_A$-GST, purified from *E.coli* grown in media supplemented with the respective metal ions, was estimated to be

5.25, 3.95 and 1.45 respectively (Tommey *et al*., 1991). The pH of half-dissociation of zinc, cadmium and copper from equine MT was determined to be 4.50, 3.00 and 1.80 respectively (Kägi and Vallee, 1960, 1961). Therefore, in comparison with equine renal MT, recombinant $PsMT_A$ protein has slightly lower affinities for zinc and cadmium, but a slightly higher affinity for copper ions.

Expression of a PCR-generated *$PsMT_A$* cDNA in *E.coli*, when grown in medium containing cadmium, resulted in an accumulation of more intracellular cadmium than the equivalent cells containing the expression vector (pPW1) alone (Kille *et al*., 1991). The effects on metal accumulation have been used as a reliable indicator of intracellular production of metal-ligand complexes in *E.coli* since expression of the wild-type plasmid (without insert) generates no such effects (cited in Kille *et al*., 1991). Metal analyses confirmed the absence of zinc and copper and indicated the estimated cadmium / $PsMT_A$ stoichiometry ranged from 5.6 to 6.1 g-atoms of cadmium per mole of protein purified from E.coli (Kille *et al*., 1991).

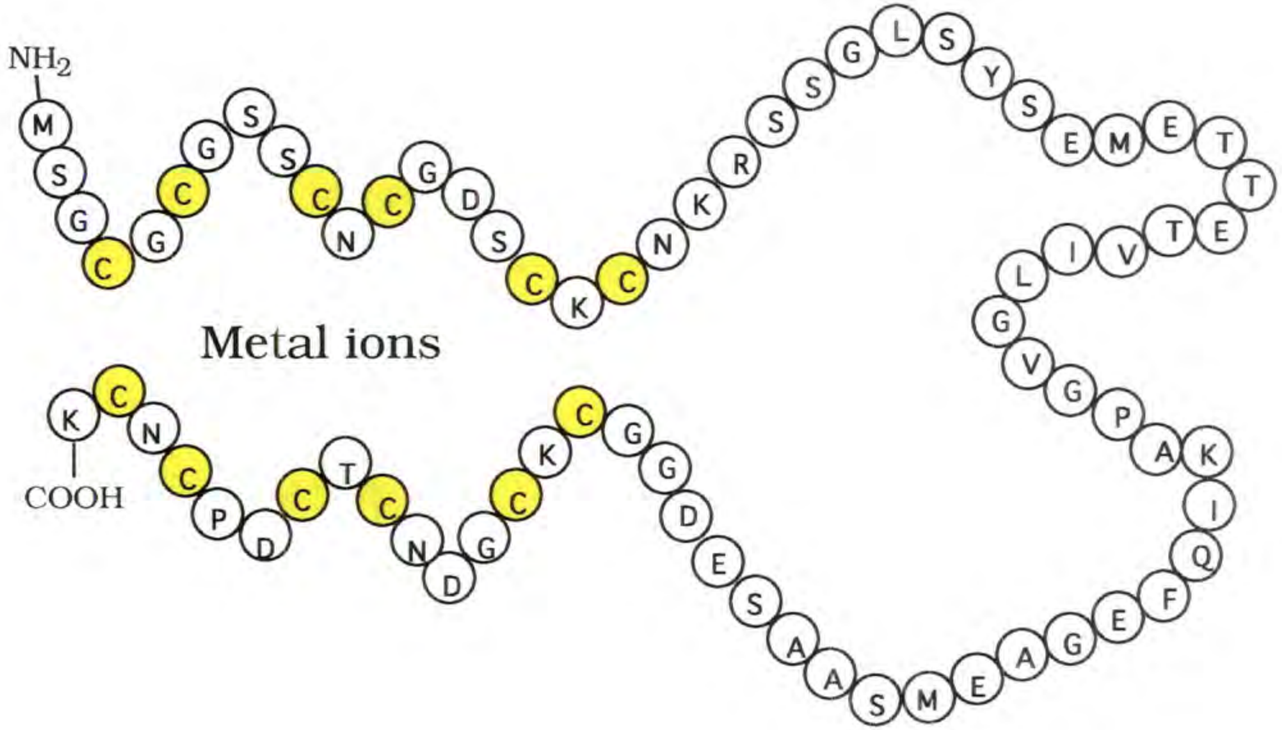
Purified recombinant $PsMT_A$ proteins were found to have undergone proteolytic degradation within the region separating the two cysteine-rich domains (Tommey *et al*., 1991; Kille *et al*., 1991). It has been demonstated that the formation of metal clusters through metal-thiolate liganding protects proteins from proteolytic degradation (Nielson and Winge, 1983). The purified protein was therefore treated with proteolytic enzymes and it was found that residues 2-21 and 56-75 in the protein are refractory to proteinase K. However, despite such cleavage within the central spacer region, the metal-liganding forces within the cysteine-rich regions appear to remain associated with each other indicating that the metal-protein bonds are capable of holding this portion of the cleaved molecule together (Kille *et al*., 1991). A putative model for metal-binding to $PsMT_A$ was proposed by Kille and co-workers (Kille *et al*., 1991) (figure 2).

Phenotypic effects of constitutive *$PsMT_A$* expression have been examined in *E.coli* cultures and *Arabidopsis thaliana* seedlings (Evans *et al*., 1992). Following

Figure 2.

A putative structure for PsMTA with metal ions.

Adapted from Kille *et al.*, 1991 and presented in Robinson *et al.*, 1993. Cysteine residues are shown in yellow.



NH_2
Metal ions
COOH

growth in media supplemented with copper, *E. coli* cells transformed with plasmid pGPMT3 and expressing $PsMT_A$ protein as a carboxyterminal extension of GST, was shown to accumulate ca. 8 times more copper than the equivalent cells containing either plasmid pGEX3X or pGPMT1. No significant effects of $PsMT_A$ expression on accumulation of zinc or cadmium were detected in cultures following growth in media supplemented with either high or low levels of these two metal ions. Evidence that the products of these MT-like genes bind metal ions within plants has been obtained from studies with *A.thaliana* transformed with a construct containing $PsMT_A$ under the control of the cauliflower mosaic virus 35 S (CaMV 35 S) promoter. In *A. thaliana* seedlings derived from a single, $PsMT_A$ -expressing F1 parent and segregating for $PsMT_A$, 75 % of seedlings accumulated more copper, up to 8-fold in the highest accumulating seedling, than untransformed control plants (Evans *et al* ., 1992).

Regulation and expression

Information describing regulation of expression of metallothionein-like plant genes is very limited. Transcripts from type 1 genes (pea, maize, barley, *Mimulus guttatus* , *Arabidopsis thaliana* and wheat) were found to be most abundant in roots. *PsMT* transcripts from pea, are not detected in the embryonic radicle but transcripts of a slightly smaller size than those in roots are detectable in the embryonic cotyledon (Evans *et al* ., 1990). The abundance of MT-L transcripts from maize is low in kernels (de Framond, 1991).

Transcripts from type 2 genes (soyabean, castorbean and *Arabidopsis thaliana*) are most abundant in leaves. In soyabean, transcripts encoding the predicted metallothionein, were also found to be present in roots but in lower adundance (Kawashima *et al.*, 1991). Related cDNAs have since been isolated from *A.thaliana* (Takahashi, 1991) and castorbean (Weig and Komor, 1992). MT2 transcripts from

A. thaliana were most abundant in leaves than in roots and dark-grown seedlings (Zhou and Goldsbrough, 1994).

Transcripts from wheatgerm E_C accumulate in immature embryos and persist in the desiccated seed but rapidly decline at germination unless supplemented with abscisic acid (ABA). This ABA-responsive element is located in the 5' flanking region of wheat E_C genes but there are no clearly identifiable metal-responsive elements on the basis of sequence similarity to metal-responsive elements known in other systems (Kawashima *et al*., 1992). The predicted product of a partial cDNA obtained from mRNA isolated from *A . thaliana* seeds (Raynal *et al*., 1993) showed sequence similarity to the E_C protein from wheatgerm.

Related cDNAs from barley (B22E) were isolated from the aleurone/pericarp and embryo of developing grains and also germinating scutella (Klemsdal *et al*.., 1991). B22E transcripts are repressed by ABA in developing seeds (Olsen *et al* ., 1990) and the corresponding gene contains consensus ABA-responsive elements (Klemsdal *et al*., 1991).

These observations raise the possibility that different members of a gene family, possibly encoding proteins with different metal-binding specificities, could be expressed in different organs or under different environmental conditions in any specific plant (cited in Robinson *et al*., 1993).

Metal regulatory elements

At present only the 5' flanking regions of two metallothionein-like genes, from pea and maize, have been characterised. Comparison of the sequences of metal regulatory elements (MREs) of animal MT genes and examination of the effect of point mutations on their function, have identified the core MRE to be 5' TGCRCNCX 3' (where R represents G or A, X represents G or C and N can be any base but A) (Stuart *et al* ., 1985; Searle *et al* ., 1985). However, in the sixth position where G, C or T give normal regulation, an A gives rise to a high level of basal expression in the absence of metals. In the 5' flanking region of pea, the sequence 5' TGCACACC 3', flanked by imperfect inverted repeats, occurs between -241 and -248 base upstream from the start site of translation. At present there is no evidence that this element is functional in pea. It was also observed, that a sequence ($PsMT_A$, 5' ATTAAGCATGCAACAAATT 3') with homology (underlined) to part of the 5' flanking region of the copper MT gene from *Neurospora crassa* (Münger *et al* ., 1987) occurs between -272 and -290 in $PsMT_A$ and part of this sequence (5' CATGCAACA 3') is repeated at -163 to -171. In contrast, no sequences in the 5' flanking region of $PsMT_A$ are significantly homologous to the control sequences of *CUP 1*, that encodes class II MT from *Saccharomyces cerevisiae* (Evans *et al* ., 1990).

The sequence of the 5' flanking region of MT-L gene from maize was searched for putative MREs, but contrary to the pea gene, no such consensus sequence was found to be present in the maize gene promoter (de Frammond, 1991).

Abscisic acid-responsive element

The E_C gene isolated from a wheat genomic library and designated gf-3.1 (*vide supra*) (Kawashima *et al* ., 1992) was found to encode an intronless mRNA for the E_C protein with appreciable amounts of 5' and 3' flanking sequences. In addition to a putative TATA box (-30), two inverted-repeat sequences (-428, -280; -408, -372) and one direct-repeat sequence (-147, -94), the 5' flanking sequence in gf-3.1 also contains a sequence similar to the abscisic-acid-responsive element (ABA -responsive elements) in other higher plant genes but does not contain sequences similar to the metal-responsive elements in animal MT genes. Comparison of the sequences of known ABA -responsive elements in wheat (Guiltinan *et al* ., 1990), rice (Mundy *et el* ., 1990) and cotton (Baker *et al* ., 1988) genes identified 5' CACGTGGC 3' as an ABA -responsive element.

Compatible with the above responsive element, a sequence -608 to -601 5'CACGTGGA 3' was identified in the 5' flanking region of gf-3.1 and is presumed to confer responses to the phytohormone.

The development of PCR methods for chromosome crawling.

PCR has rapidly become establised as a powerful technique for both gene analysis and cloning (Saiki *et al.*, 1988). The specificity of PCR amplification is normally based on two oligonucleotide primers of known sequence that flank the DNA segment to be amplified and which hybridise to opposite strands. The primers are oriented so that DNA synthesis proceeds across the specific region between the primers (Erlich *et al* ., 1988).

PCR was originally used with single primer species. However, it is now possible to obtain specific DNA amplification using redundant primers derived from an amino acid sequence where only a limited portion of a protein sequence is known (Lee *et al* ., 1988; Gould *et al* ., 1989). The primers should be selected from a sequence that exhibits minimal redundancy since a lower degree of primer degeneracy reduces the possibility of mis-priming and non-specific amplification events. The use of peptide regions containing amino acids encoded by four or six codons should therefore be avoided. Fortunately, degenerate PCR primers based on as few as three consecutive amino acids can be sucessfully used by including 6-9 base 5' extensions, usually incorporating a restriction enzyme site. Although not complementary to the template, these 5' extensions become incorporated into the amplified product during the second cycle and all subsequent cycles of amplification. This improves the efficiency of amplification due to the increased stability of the priming duplex (Mack and Sninsky, 1988). Alternatively, the neutral base analogue, inosine can be incorporated into primers corresponding to polypeptide regions which would otherwise require complex oligonucleotide mixtures (Fordham-Skelton *et al*., 1990).

The original protocols for PCR (Saiki *et al*., 1985; Mullis *et al*., 1986; Mullis and Faloona, 1987) used the Klenow fragment of DNA Polymerase I from *Escherchia coli* (*E.coli*), to catalyse the extension of the annealed primers. However, the thermolabile Klenow fragment is inactivated by the heat denaturation step, used to separate the newly synthesised strands and a fresh aliquot of enzyme is required during each successive cycle. This becomes a tedious and error prone process if several samples are amplified simultaneously. Reactions using this enzyme generally worked well for the amplification of small segments of genomic DNA or cDNA (<200 bp in length). The results with larger templates were poor, in terms of product yield and products were often heterogeneous in size (Sharf *et al*., 1986). In addition, mis-priming was often observed due to the low temperatures (37 °C) required for the catalytic extension of the primers.

To resolve these problems, a 94 kd thermostable DNA polymerase (*Taq* polymerase) was isolated from the thermophilic bacterium *Thermus aquaticus* (Chien *et al*., 1976). This was used as an alternative to DNA polymerase I from *E.coli* (Saiki *et al*., 1988). Due to its thermostability the enzyme from *T. aquaticus* can survive extended incubation at temperatures as high as 95 °C, eliminating the need for it to be replenished at the beginning of each cycle. Furthermore, by enabling the extension reaction to be performed at higher temperatures (72 °C), within the optimal temperature range of this particular polymerase (70-80 °C), significantly improves the specificity, yield and sensitivity of the reaction. Therefore, single-copy genomic sequences may be amplified by a factor of more than 10 million with very high specificity, and 0.5-1.0 μg DNA segments up to 2 kb in length may be amplified (Saiki *et al*., 1988) from 30-35 cycles. Eventually, the accumulation of product begins to limit the reaction as there is insufficient enzyme present to extend all the primer-template duplexes. When this occurs, product accumulates in a linear rather than exponential manner. If further amplification of the target sequence is required, a sample of the amplified DNA can be diluted 1000- to 10 000- fold and used as the template for subsequent rounds of synthesis in a fresh PCR reaction. Using this method, amplification levels of 10^{10} can be achieved during 60 sequential cycles (Saiki *et al.*, 1988). This allows detection by southern hybridisation of a single copy of the target sequence in the presence of a 10^{13} fold excess of non-target DNA.

Although the polymerase chain reaction is a powerful technique, there are drawbacks associated with its use. PCR only requires a single copy of a gene as a template for amplification, therefore, meticulous care must be taken to avoid sample contamination with other DNA and hence prevent synthesis of "false positives". Carryover of DNA from a previous amplification of the same target sequence is the major contamination problem. This is especially the case when amplifying gene homologues from different species or amplifying segments of DNA using the same primers which have previously been used to amplify cloned products. Using the same principles applied to sterile handling of cell cultures cross-contamination may be minimized. It is, therefore, necessary to set up positive and negative controls and to follow stringent procedures to minimize PCR carryover (Kwok, 1989; Cimino *et al* ., 1990; Kitchen *et al.* 1990; Sarkar and Sommer, 1990).

One problem associated with PCR is that DNA polymerase make errors, amplifies them and introduces these errors into each new cycle. For *Taq* polymerase, A.T to G.C transitions, occur at a frequency of 1 in 400 accumulated mutations after the usual 20-30 cycles (Saiki *et al.*, 1988). Although the frequency of PCR errors is less when using Klenow (1 in 600 accumulated mutations), the disadvantages of using Klenow are well documented (*vide supra*). The amplification of DNA *in vitro* is five to seven orders of magnitude less accurate than replications of plasmids and phages in *E.coli* (Karlovsky, 1990). The nucleotide sequence of a PCR clone should therefore only be regarded as a confirmation of its identity or as a probe for screening (Fordham-Skelton *et al* ., 1990). If PCR products are to be cloned and then sequenced *Taq* polymerase may be replaced by *T4* polymerase (Keohavong *et al,* 1988), which makes less errors due to its proof-reading activity. It is, however, important that a number of independent clones from independent PCR reactions are analysed to obtain the correct sequence.

Methods for Chromosome Crawling.

A significant limitation of conventional PCR is that DNA sequences situated outside the primers are inaccessible. An oligonucleotide that primes synthesis into a flanking region, would produce only a linear increase in the number of copies rather than an exponential increase which occurs when there is an oligonucleotide in the reverse direction.

Two methods have been developed which allow the *in vitro* amplification of DNA regions flanking a known sequence: These techniques have been termed Inverse polymerase chain reaction and Ligation-mediated or anchored polymerase chain reaction.

Inverse Polymerase Chain Reaction (IPCR).

This method was developed independently by three groups (Ochman *et al* ., 1988; Tirglia *et al* ., 1988; Silver and Keerikatte, 1989) and is based on digestion of genomic DNA with a restiction enzyme that has no cleavage sites within the target sequence. In addition the fragment carrying the target sequence should be no greater than 2-3 kb in length. The DNA is diluted and ligated under conditions which favour the formation of monomeric circles. *Taq* DNA polymerase works slightly more efficiently with linear DNA than with circular DNA and consequently some workers (Triglia *et al* ., 1988) relinearise the template by digesting with a restriction enzyme that cleaves once within the known sequence of the target DNA. Whether linear or circular templates are used, the regions flanking the target sequence can be amplified in a polymerase chain reaction using primers which will anneal in the opposite orientation to those normally employed for PCR. The major amplification product of the reaction is, therefore, a linear double-stranded DNA molecule which consists of a head-to-tail arrangement of sequences flanking the original target DNA. The junction between the upstream and downstream sequences is marked by the presence of a restriction site for the enzyme originally used to digest the genomic DNA.

Ligation-Mediated or Anchored Polymerase Chain Reaction (APCR).

Ligation-mediated or anchored PCR was developed by Shyamala and Ames (1989) and is an alternative technique to inverse PCR for chromosome crawling. Anchored PCR allows amplification of unknown DNA sequences, when sequence information is only available at one position. Thus, providing a short length of DNA sequence to design a gene-specific primer.

APCR is dependent upon the ligation of restricted genomic DNA fragments to a sequence which contains a generic primer binding site (eg M13 primer binding sites in pUC or pBluescript cloning vectors), instead of producing circular ligation products as required for IPCR.

A variety of fragments will be ligated to the vector DNA. However, during amplification the vector-specific primer will anneal to all complexes but only the gene-specific primer will anneal to the target sequence. This fragment will accumulate exponentially, whereas the non-specific fragments will only accumulate at a linear rate, as only the vector-specific primer binding site being available during PCR amplification.

Shyamala and Ames (1989) were able to take two contiguous steps using ligation-mediated PCR with single-specific primer species into the region flanking of the known histidine transport operon of *Salmonella typhimurium*.

Robinson and co-workers (1990) combined and adapted several of these methods to amplify and clone a prokaryotic metallothionein gene, designated *smt* A, from the cyanobacterial strain *Synecochoccus* PCC6301. This strategy employed inosine-containing oligonucleotide primers for both PCR and APCR. The primers were derived from known protein sequence and used to amplify the coding region of *smtA* and subsequently "chromosome crawl" into the 5' and 3' flanking regions of the gene. This strategy has general applications for the rapid characterisation of novel prokaryotic genes when only limited amino acid sequence is available (Robinson *et al* ., 1990). However, this strategy has not, as yet, been successfully employed in plants.

The development of DNA sequencing technology.

Two methods for obtaining DNA sequence data are by chemical cleavage (Maxim and Gilbert, 1977) and the most commonly used enzymatic chain-terminating dideoxynucleotide (Sanger *et al* ., 1977). Both of these methods require relatively large amounts of sequencing template, incorporating radiolabelled nucleotides and exposure to X-ray film. These methods are very time-consuming and labour-intensive. However, in recent years, the Sanger dideoxynucleotide chain-terminating method has undergone a number of specific modifications: Radioactive labelling has been replaced by fluorescent labelling, either utilising a one dye or four dye labelling system; the klenow enzyme has been replaced by *Taq* polymerase enabling extension reactions to be performed at higher temperatures, reducing most secondary structure problems; the deoxyguanosine triphosphate (dGTP) has been replaced by deaza-guanosine triphosphate (c^7dGTP) to minimise band compressions; and a new method for performing the reactions on a DNA thermal cycler has been developed and is termed 'cycle sequencing'.

Automation of DNA sequencing.

The repetitious nature of DNA sequencing makes it suitable for full or partial automation. Several commercial instruments are currently available for automated analysis of the DNA ladders generated in the sequencing reactions, based on either fluorescent primers or terminators (Smith *et al*., 1985; Smith *et al*., 1986; Ansarge *et al*., 1986; Prober *et al*., 1987; Kambara *et al*., 1988). During electrophoresis, an argon laser, which is either fixed or scans across a line near the bottom of the gel, is used to excite fluorescence from the passing dye-labelled DNA products. The emitted light is filtered, collected, and transformed into digital signals that are automatically analysed by a computer to interpret the DNA sequences. However, these instruments only automate the 'backend' part of the analysis, ie. the electrophoretic separation of DNA fragments and the conversion of this information into nucleotide sequences. Complete automation would involve the 'frontend' (ie. the sequencing reactions) being automated. This has been made possible by using any of several available programmable pipetting robots for full automation of the sequencing reactions or semi-automation using DNA thermal cyclers, leaving only the selection of templates and gel loading to be manually performed.

Methods for template preparations have also changed significantly, which will be discussed further in a later chapter.

Cycle sequencing.

Although cycle sequencing resembles PCR, amplification is only from a single strand. PCR amplification, on the other hand is exponential since two strands are synthesised simultaneously, as previously described.

Cycle sequencing reactions are performed using a DNA thermal cycler. Compared with conventional sequencing procedures, thermal cycling of the sequencing reactions increases signal intensity and decreases sensitivity to reaction conditions.

Using a thermal cycler removes many variables that often cause problems in standard bench sequencing.

During cycle sequencing, DNA synthesis requires a template that is in a single-stranded conformation. This is achieved by heating the sample in solution, disrupting the hydrogen bonds between the complementary bases of the two strands. Primer annealing can occur when the temperature is brought down to a level below its melting temperature (Tm). *Taq* polymerase is then used for primer extension. This reaction process is then repeated 25 or 30 cycles. The fluorescent signal can be amplified to levels which allow enhanced base calling by an automated DNA Sequencer and analysis software.

There are a number of advantages for cycle sequencing: no alkaline denaturation is required for ds templates; starting material and enzyme required is considerably less; the same protocol is used for ss and ds DNA and PCR-generated fragments; longer lengths of DNA can be read; there is less "hands on" work required; and potentially larger constructs (in the order of 40kb) can be sequenced.

Fluorescent DNA Sequencing Chemistries.

The biochemistry of the four-dye, one lane sequencing system is based on the Sanger chain-termination method and has been developed by Applied Biosystems. Four fluorescent dyes (green, blue, yellow and red), are associated with each one of the four dideoxynucleotide (A, C, G and T respectively). The four-dye system allows multiplexing the reactions in one tube and the products analysed together in a single electrophoresis gel lane. Thus, reducing any lane to lane variation possibly observed by other one-dye, four lane systems. There are two different labelling systems: Dye-labelled primers and dye-labelled terminators.

Dye - labelled primers.

The four fluorescent dyes used for dye-primer sequencing are of two structural types; Joe and Fam (A + C) molecules are fluorescein dyes while Tamra and Rox (G + T) molecules are rhodamine dyes. The fluorescent dyes appear to interact exclusively with the first five bases on the 5' end of the oligonucleotide to which they are attached. It is this interaction which is responsible for their effect on primer mobility, ie their different sizes and the manner in which they interact with the bases of the 5' end of the oligonucleotide, imparts an effect on the electrophoretic mobility of terminated fragments. This chemistry utilises four tubes for performing the reactions but are combined into a single tube prior to electrophoresis.

Dye - labelled terminators.

Fluorescent DNA sequencing may also be performed using a chemistry in which the fluorescent dyes are attached to the dideoxynucleotides (ddNTPs). The fluorescent molecules attached to the ddNTPs are different to those used in dye-labelled primers, ie they are all of the rhodamine type. Due to the similarity of the dyes, they affect fragment mobilities equivalently, so no mobility shift correction is applied in the base calling analysis programme for dye-labelled terminators. However, the colours of the dyes used for dye-labelled terminators do not match those used for dye-labelled primers. Consequently, for the chromatographic display of the analysed data, the analysis program reverses the dye-labelled terminator colours to those conventional for dye-labelled primers. This chemistry utilises a single tube for performing the reaction but requires a purification step to remove unincorporated dideoxynucleotides prior to electrophoresis.

Sequencing DNA fragments generated by the polymerase chain reaction.

A significant use of the polymerase chain reaction is for the generation of sequencing templates, from either cloned inserts or directly from genomic DNA, for use in a variety of clinical and basic research situations.

Traditionally, these sequencing templates have been obtained by inserting the target DNA sequence into bacterial or viral vectors which are then multiplied in bacterial host cells. These cloning methods have been simplified and standardised but problems remain with the maintenance and use of living cell system, (such as *de novo* mutations in vector and host cell genomes). The use of PCR allows the generation of sequencing templates more efficiently than with the present cell dependent methods. Amplification of cloned inserts of unknown sequence can be achieved by using oligonucleotides that prime inside, or close to the polylinker of the cloning vector (Saiki *et al*., 1988). Thus, there no longer exists a need for repetitive growing and purification of sequencing templates from small cultures of bacteria or viruses.

Sequencing PCR products directly has three advantages over sequencing cloned PCR products: 1) sequence analysis is readily standardised since it is a simple enzymatic process, 2) only a single sequence needs to be determined for each sample and 3) reduces the time normally incurred for sub-cloning. In contrast, a consensus sequence based on several cloned PCR products has to be determined for each sample, in order to distinguish any mutations present in the original genomic sequence from random mis-incorporated nucleotides, introduced by *Taq* polymerase during PCR (Saiki *et al.* 1988).

Methods for improving PCR templates for direct sequencing.

The simplicity of the PCR principle and the general robust nature of the reaction in generating ds DNA products, contrasts with the difficulty encountered when attempting to sequence the amplified fragments directly.

Problems associated with direct sequencing of PCR products may not be due to lack of specificity, but result from the ability of the two strands of the linear amplified product, to rapidly reassociate after denaturation. This re-naturation either blocks the primer-template complex from extending or prevents the sequencing oligonucleotide from annealing efficiently (Gyllensten and Erlich, 1988). Other problems may arise from products being heterogeneous or there is insufficient template required for the sequencing reaction.

A number of methods have been used to improve template quality for both ss DNA and ds DNA PCR products for sequencing. However, some methods for purification of PCR products are inconvenient, time-consuming and inefficient. In the course of these studies several different methods for obtaining sequences of PCR -generated products suitable for sequencing have been evaluated and optimised. These include the following:

The PCR products may be purified by:

1. transferring the product to DEAE cellulose, a modification of the procedure described in Sambrook *et al.*. (1989) for the purification of size-fractionated DNA fragments directly from agarose gels (J.V. Hookey, personal communication);
2. a modification by Zhen and Swank (1993) of the procedure developed by Hogness described in Sambrook *et al.*. (1989) for the purification of size-fractionated DNA fragments directly from agarose gels has been developed and adapted for PCR products (J.V. Hookey, personal communication). The products are purified in the presence of polyethylene glycol (PEG).
3. for relatively fast procedures for obtaining purified PCR products commercially available kits have been used. These include Promega Magic^{TM} PCR Preps DNA Purification System for Rapid Purification of DNA Fragments and QIAGEN QIAquick-spin PCR Purification Kit.

Sequencing Plasmid and Cosmid DNA.

There are a number of problems associated with sequencing plasmid and cosmid DNA. Cycle sequencing relies on accurate quantification and sample purity, as the result of any DNA sequencing process is directly proportional to the quality and quantity of the sample preparation. In the course of these studies several methods have been evaluated and optimised for sequencing and templates are prepared using one of the proven methods (refer to methods for preparing DNA for sequencing).

The Aims of this Research.

MT genes and their translational products have been identified and characterised from animals, fungi and cyanobacteria, but the precise function (s) of most of these molecules is/are not yet fully understood. Metallothionein-like genes have been identified from several plant species but at present their translational products remain to be purified from plant material and sequenced. The wheat E_C protein (Lane *et al* ., 1987) is the only plant protein which can be unequivocally designated a metallothionein since its product has been purified. An understanding of the mechanisms controlling the expression of these genes is likely to contribute towards better understanding of their functions.

MT gene expression is rapidly and transiently induced by exposure to trace metals, and in mammalian cells, various circulating factors such as hormones and interferon are also known to act as inducers. The *cis* -acting control sequences of MT genes in a number of organisms have been studied and *trans* -acting factors with which they interact, in particular to confer metallo-regulation, have been identified. However, information describing the regulation of expression of metallothionein-like plant genes is very limited.

It has been hypothesised that plants may contain a family of MT -like genes (refer to figure 1) whose products could perform different functions in the metabolism / detoxification of different metal ions. Different members of these gene families may be subject to different environmental and developmental controls, in consequence, these genes may contain differing complements of *cis* -acting elements. For example, *A. thaliana* has at least three MT -like genes (Takahashi, 1991; Zhou and Goldsbrough, 1993; Raynal *et al* ., 1993) possibly encoding proteins with different metal-binding specificities, which could be expressed in different organs or under different environmental conditions.

Transcripts have been identified within different tissues of several plants. To date, proposed type 1 transcripts have been found to be most abundant in roots of several species, type 2 transcripts most abundant in leaves and a possible third type referred to as E_C most abundant in seeds. At present only the 5' flanking regions of two type 1 MT -like genes, from pea (Evans *et al*., 1990) and maize (de Framond, 1991), have been reported and no homologous regions were identified. However, it was observed that a 19 bp sequence in the 5' flanking region of $PsMT_A$ from pea showed some homology to part of the 5' flanking region of the copper responsive MT gene from *N. crassa* (Evans *et al*., 1990).

The metal-binding characteristics of the products of the MT -like gene from pea suggest that these proteins (at least those designated type 1) bind strongly to copper and may have a role in the homeostasis of copper. Expression of the $PsMT_A$ gene is also responsive to iron, as is another type 1 sequence, *ids1*, in barley (Okumura et al., 1991). However, it has been suggested that this may be an indirect response mediated by changes in internal copper coincident with changes in exogenous iron. These type 1 genes may therefore be predicted to have *cis* -acting copper responsive elements and possibly regulatory elements which mediate expression in roots. Type 2 sequences may be predicted to contain *cis* -acting elements which confer expression in leaves and possibly responses to metals. Finally, E_c type genes do not appear to be metal responsive but must contain elements which confer developmental control of expression in seeds. Two 5' flanking regions have been reported for E_c type genes and consensus ABA -reponsive elements have been identified within these sequences (Kawashima *et al*., 1991; Klemsdal *et al*., 1991).

Southern analyses using $PsMT_A$ as a probe suggested a small multigene family from pea and putative homologues from *P. vulgaris*, rape and tobacco. The aim of this research was, therefore, to determine the sequences of the 5' flanking regions of several different MT -like genes from several plants. This will allow any consensus sequences (within a particular group of MT -like genes), which could be regulatory elements that control the expression of these genes, to be identified.

In conducting these studies it is intended to further apply novel polymerase chain reaction techniques (ligation-mediated PCR) or as an alternative, screen genomic libraries to facilitate the isolation of gene flanking regions. It is also intended to optimise automated

DNA sequence analysis with respect to DNA template preparation and the direct sequencing of PCR products and cosmid DNA.

CHAPTER 2
MATERIALS AND METHODS

MATERIALS

Chemicals and equipment.

All chemicals and biological reagents, with the exception of those noted below, were from BDH-Merck Ltd and were of analytical grade.

Caesium Chloride, Glycogen, *Taq* polymerase, *Taq* polymerase buffer, Deoxy-nucleosides-triphophates set, Restriction and modifying enzymes were supplied by Boehringer Mannheim UK Ltd., Lewes, East Sussex.

Acrylamides, Urea, Antibiotics, Ethidium Bromide, Herring sperm DNA, RNase, Mineral Oil light, Bovine Serum Albumin, Polyvinyl-pyrrolidone, N-Z-Amine A, Formamide and Amberlite MB-1 resin were supplied by Sigma Chemical Co. Ltd., Poole, Dorset.

Alconox™ Detergent was supplied by Aldrich Chemical Co. Ltd., Gillingham, Dorset.

Agarose, DH5α competent cells, Bacterial strain HB101, pUC 18 vector were supplied by Gibco BRL, Life Technologies Ltd., Paisley, Scotland.

Ficoll, Minigel apparatus, Gene ATAQ Controller, 40-well, were supplied by Pharmacia Biosystems, Milton Keynes.

Bacterial strains LE392 and KW251, pGEM T vector, pGEM 3Z vector, Magic and Wizard miniprep Kits were supplied by Promega Ltd., Southampton.

pBluescript SK+ and KS+ plasmids were supplied by Stratagene Ltd., Cambridge.

DNA size markers, IPTG, X-Gal, Restriction and modifying enzymes were supplied by Northumbria Biologicals Ltd., Cramlington, Northumberland.

Bacto-tryptone was supplied by Becton Dickinson, Cowley Oxon.

Bacto-agar and Casamino Acid were supplied by Difco Ltd., W. Molesey, Surrey.

Oxoid Yeast extract was supplied by Unipath Ltd., Basingstoke, Hampshire.

Petri dishes were supplied by Bibby Sterilin Ltd., Stone, Staffordshire.

Filter units, Acrodisc, were supplied by Gelman Sciences Ltd., Northhampton.

Membrane filters (0.2μm), 3MM filter paper were supplied by Whatman International Ltd., Maidstone, Kent.

Nitrocellulose filters, Schleicher and Schuell grade BA-85, was supplied by Anderman & Co. Ltd., Kingston-upon Thames.

Radiochemicals were supplied by Amersham International plc., Aylesbury, Buckinghamshire.

National Diagnostics "Ecoscint", scintillation fluid, was supplied by B.S.& S (Scotland), Edinburgh.

X-ray film, Fuji-RX, was supplied by Fuji Photo Film (UK) Ltd., London.

X-ray cassettes were supplied by Genetic Research Instrumentation Ltd., Dunmow, Essex.

Developer, Ilford phenisol, was supplied by Ilford Ltd., Mobberly, Cheshire.

Fixer, Kodac Unifix, was supplied by Phase separations Ltd., Deeside, Clwyd.

Qiagen Lambda DNA purification kit and Qiagen QIAprep-spin Plasmid purification Kit were supplied by Hybaid Ltd., Teddington, Middlesex.

Techne PHC-3 thermal cycler, 96-well, 0.25ml tubes, heat sealing film, hybridisation tubes and ovens were supplied by Techne (Cambridge) Ltd., Duxford, Cambridge.

Automated DNA Sequencer, DNA Synthesiser, chemicals related with these instruments (except chemicals for preparing sequencing gels) were supplied by Applied Biosystems Ltd, Warrington, Cheshire.

Sequagel™-6, sequencing gel mixture, was supplied by National Diagnostics Ltd., Hessie, Hull.

Ammonium Persulphate was supplied by International Biotechnologies, Inc., New Haven, Connecticut, USA.

Biospin 30 chromatography columns were supplied by Bio-Rad Laboratories Ltd., Hemel Hempstead, Hertfordshire.

Commonly used buffers, media and other solutions.

All glassware and plasticware were washed thoroughly with alconox™ detergent and rinsed several times with distilled water. All buffers and solutions were prepared using double deionised water (milli Q) and sterilised by autoclaving (20 minutes at 120 °C) or ultrafiltration (0.2 µm) as appropriate.

LB medium (per litre)	10 g Bacto-tryptone
	5 g Yeast extract
	5 g NaCl

LB agar as above + 15 g l^{-1} Bacto-agar

NZCYM media (per litre)	10 g N-Z-Amine A from bovine milk
	5 g NaCl
	5 g Yeast extract
	1 g Casamino acids
	2 g $MgSO_4$
	15 g Agarose
	Adjust pH to 7.5 with NaOH

NZCYM top agarose as above + 7 g l^{-1} Agarose

SOC medium	2 % Bacto-tryptone
	0.5 % Yeast extract
	10 mM NaCl
	2.5 mM KCl
	10 mM $MgCl_2$
	10 mM $MgSO_4$
	20 mM Glucose

Antibiotics were added to media and agar plates as required:

Ampicillin (70 % ethanol)	100 μg ml^{-1} final concentration
Carbenicillin (water)	100 μg ml^{-1} final concentration
Tetracycline (DMSO)	10 μg ml^{-1} final concentration

X-gal (DMF) was added to agar plates as required:

40 μg ml^{-1} final concentration

IPTG (water) was added to agar plates as required:

100 mM final concentration

Phage buffer	20 mM Tris-HCl pH 7.4
	100 mM NaCl
	10 mM $MgSO_4$
SM buffer	20 mM Tris-HCl pH 7.4
	100 mM NaCl
	10 mM $MgSO_4$
	0.01 % Gelatin
TAE buffer	40 mM Tris-acetate pH 7.7
	10 mM EDTA
TBE buffer	89 mM Tris pH 8.3
	89 mM Boric Acid
	2 mM EDTA

Denaturing solution	1.5 M NaCl 0.5 M NaOH 1 mM EDTA
Neutralising solution	3 M NaCl 0.5 M Tris-HCl pH 7.0 1 mM EDTA
20x SSC	3 M NaCl 0.3 M Tri-sodium citrate pH 7.0 with HCl
Denhardt's solution	0.02 % Ficoll 400 0.02 % Polyvinyl-pyrrolidone 0.02 % BSA
TE buffer	10 mM Tris-HCl pH 7.5 1 mM EDTA
Agarose gel loading dye	0.1 % Orange G 20 % Ficoll 400 0.1 M EDTA pH 8.0
Sequencing gel loading buffer	1 part 50 mM EDTA pH 8.0 5 parts deionised formamide

Phenol was redistilled, saturated with TE buffer and stored frozen at -20 °C. Formamide was deionised by stirring for 30 minutes with amberlite MB-1 resin, then filtered through Whatman 0.2 μm membrane filters and stored at -20 °C in 1 ml aliquots.

Sodium Iodide (NaI) solution was prepared by adding 90.8 g NaI and 1.5 g Na_2SO_3 to 100 ml milli Q water (the final volume is greater than 100 ml). The solution was filter sterilised and a further 0.5 g Na_2SO_3 added to produce a saturated solution. The resulting solution was stored in the dark at 4 °C.

Bacterial Strains.

The genotypes of the bacterial strains used are:

JM101

sup E, *thi*, Δ (*lac pro* AB), [F', *tra* D36, *pro* AB, *lac* I^qzM15].

JM109

end A1, *rec* A1, *gyr* A96, *thi*, *hsd* R17 (r_K^-, m_K^+), *rel* A1, *sup* E44, Δ (*lac pro* AB), [F', *tra* D36, *pro* AB, *lac* I^qZΔM15].

DH5a

F-, 80d*lac* ZΔ M15, *rec* A1, *end* A1, *gyr* A96, *thi* -1, *hsd* R17 (r_K^-, m_K^+), *sup* E44, *rel* A1, deo R, Δ(*lac* ZYA-*arg* F)U169.

KW251

F-, *sup* E44, *gal* k2, *gal* T22, *met* B1, *hsd* R2, *mcr* B1, *mcr* A, [*arg* A81:Tn10], *rec* D1014.

LE392

F^-, *hsd* R574, (r_K^-, m_K^+), *sup* E44, *sup* F58, *lac* Y1 or Δ(*lac* IZY)6, *gal* k2, *gal* T22, *met* B1, *trp* R55.

HB101

F^-, *hsd* S20, (r_B^-, m_B^-), *sup* E44, *rec* A13, *ara* - 14, *leu* B6, *pro* A2, *lac* Y1, *rps* L20 (str^r), *xyl* -5, *mtl* -1.

Pea genomic library (DPB) was a gift from Dr D.P. Bown. Pea genomic library (AMT), *Arabidopsis thaliana* cDNA and genomic libraries were a gift from Dr A.M. Tommey.

METHODS

GENERAL MOLECULAR BIOLOGY PROCEDURES

Unless otherwise stated all general procedures were performed as described by Sambrook *et al.* (1989).

Isolation of plasmid, cosmid and phage DNA.

Small scale preparation of plasmid and cosmid DNA for restriction analysis.

A 10 ml overnight culture of the bacteria was grown with the appropriate antibiotic selection. Cells (3 ml) was centrifuged for 5 minutes at 6,000 x g in a microcentrifuge. The supernatant was removed and the pellet partially dried by inversion of the tube over tissue paper, 100 μl of ice cold solution 1 (50 mM glucose, 10 mM EDTA pH 8.0, 25 mM Tris.Cl pH 8.0) was added. Cells were resuspended using a pipette and incubated at room temperature for 5 minutes. An aliquot (200 μl) of freshly prepared solution 2 (0.2 N NaOH, 1.0 % SDS) was added, mixed gently by inversion and incubated on ice for 5 minutes. An aliquot (150 μl) of ice cold solution 3 (11.5 ml glacial acetic acid, 28.5 ml water, 60 ml 5 M potassium acetate) was added, mixed by inversion and incubated on ice for 5 minutes. Chromosomal DNA and other cellular debris were pelleted by centrifugation at 12,000 x g for 3 minutes. The supernatant was transferred to a fresh tube, centrifuged for a further 3 minutes at 12,000 x g to obtain a particle free supernatant and transferred to a fresh tube. RNase A was added to a final concentration of 20 μg ml^{-1} and incubated at 37 °C for 20 minutes. An equal volume of phenol/chloroform (1:1) was added to the RNase A treated supernatant, mixed thoroughly for 30 seconds using a whirlimixer and centrifuged for 1 minute at 12,000 x

g. The aqueous phase was tranferred to a fresh tube, precipitated with 3 volumes of 100 % ethanol and DNA precipitated at -20 °C for 2 hours or -80 °C for 20-30 minutes. DNA was pelleted by centrifugation at 12,000 x g. The pellet was washed with 70 % ethanol, air dried and resuspended in 50 µl H_20 or TE buffer pH 7.5. An aliquot (1 µl) was digested with specific restriction endonucleases and separated by agarose gel electrophoresis.

Large scale preparation of cosmid DNA.

LB-media (10 ml) containing tetracycline was inoculated with a loop full of freshly streaked out bacteria and incubated overnight at 37 °C. This 10 ml culture was added to 500 ml of LB containing tetracycline and incubated overnight at 37 °C with vigorous aeration. Cells were harvested by centrifugation at 4000 x g for 15 minutes in a Beckman centrifuge (JA-10 rotor head, 6 x 250 ml centrifuge bottles), the supernatants discarded and the bacterial pellets placed onto ice. Each pellet was resuspended in 10 ml solution 1 (50 mM Glucose, 25 mM Tris-Cl pH 8.0, 10 mM EDTA) using a pipette and transferred into two 50 ml Oak Ridge centrifuge tubes. An aliquot (10 ml) of solution 2 (0.2 N NaOH, 1 % SDS) was added to each tube, mixed by inversion and incubated on ice for 10 minutes. An aliquot (7.5 ml) of ice cold solution 3 (11.5 ml glacial acetic acid, 28.5 ml water, 60 ml 5M potassium acetate) was added to each tube, mixed by vigorously inverting the tubes several times and incubated on ice for 10 minutes. Chromosomal DNA and other cellular debris were pelleted by centrifugation at 15,000 x g for 15 minutes in a Beckman centrifuge (JA-20 rotor head, 8 x 50 ml centrifuge bottles), the supernatant was filtered through miracloth into a beaker and transferred into two 50 ml Oak Ridge centrifuge tubes. Isopropanol was added up to the neck of the tubes (at least 0.6 volume), mixed by inversion and incubated at room temperature for 15 minutes and the precipitate collected by centrifugation at 10,000 x g for 10 minutes in a Beckman centrifuge (JA-20 rotor head, 8 x 50 ml centrifuge bottles). Pellets were washed with 70 % ethanol and partially dried in a vacuum dessicator. TE buffer (4 ml) was added to each tube to resuspend the pellets, combined into one tube and 8.6 g of caesium chloride (CsCl) added. Once the CsCl had dissolved, 0.45 ml of ethidium bromide (10 mg ml^{-1} stock) was added and the solution placed into two Beckman quick seal centrifuge tubes (0.5 x 2 inch). Tubes were filled to a level 1/16 of an inch into the neck using a CsCl solution (8 ml H_2O, 8.6 g CsCl), balanced and the tubes heat sealed. The sealed tubes were placed in a Sorvall ultracentrifuge (Type 70 angle head rotor), overnight (16 hours) at 50,000 x g at 15 °C. After centrifugation the tubes were examined under U.V. illumination., 2 bands were visible, the upper band contained chromosomal DNA, the lower band corresponded to cosmid DNA. Using a needle and syringe, the lower band was recovered and the mixture extracted several times with an equal volume of butan-1-ol saturated with CsCl to remove all the ethidium bromide. Once the sample was free of ethidium bromide, it was dialysed against TE buffer overnight at 4 °C. The sample was transferred to a 50 ml corex tube, DNA precipitated by addition of 1/5 volume ammonium acetate, 3 volumes of cold 100 % ethanol and incubated overnight at -20 °C. DNA was collected by centrifugation at 12,000 x g for 15 minutes in a Beckman centrifuge (JA-20 rotor head, 8 x 50 ml centrifuge bottles), washed with 70 % ethanol, air dried and resuspended in 1 ml sterile milli Q water. The concentration and purity of the sample was determined by absorbance spectrophotometry (260 nm and 280 nm) and agarose gel electrophoresis.

Small scale preparation of phage lambda DNA.

The method to obtain phage lambda DNA was an adaptation of the plate lysate procedure described by Sambrook *et al* . (1989).

Single hybridising phage plaques were removed from agar plates with the thin end of a pasteur pipette and placed in 1 ml of SM buffer in Eppendorf tubes. The tubes were left at room temperature for at least 2 hours to allow the phage to diffuse out from the agar plug and a phage titre had to be established in order to obtain confluent lysis to isolate phage DNA from plate lysates. A host bacterial strain was grown overnight at 37 ^{0}C in LB containing 0.2 % maltose and 10 mM $MgCl_2$. Cells were collected by centrifugation at 4,000 x g, resuspended in 0.4 volume 10 mM $MgCl_2$ and stored at 4 ^{0}C until required. An aliquot (200 µl) of previously prepared cells was added to 3 ml NZCYM top agarose (50 °C), poured over NZCYM agarose plates and air dried for 15 minutes. Aliquots (50 µl) of diluted phage (undiluted - 10^{-5}), were spotted onto the plates, air dried and incubated overnight at 37 ^{0}C. Once a suitable dilution had been established to obtain confluent lysis, 200 µl of the phage dilution was absorbed to an equal volume of a host bacterial strain at 37 ^{0}C for 20 minutes. The phage and cells were mixed with 3 ml of NZCYM top agarose at 50 ^{0}C, poured over prewarmed (37 ^{0}C) NZCYM agarose and incubated overnight at 37 ^{0}C until confluent lysis was visible. SM buffer (4 ml) was added to the plate and rotated slowly on a rocking platform to allow the phage to elute for at least 2 hours. The phage-SM suspension was transferred to a 15 ml corex tube, the plate was rinsed with 1 ml SM and transferred to the corex tube. Bacterial debris was removed by centrifugation at 8,000 x g for 10 minutes at 4 ^{0}C. The supernatant was transferred to a fresh 15 ml corex tubes, RNase A and DNase 1 were added to a final concentration of 1 µg ml^{-1} and incubated at 37 ^{0}C for 30 minutes. An equal volume of 20 % PEG-8000, 2 M NaCl prepared in SM buffer was added and incubated for 1 hour at 0 ^{0}C. Precipitated phage particles were collected by centrifugation at 10,000 x g for 20 minutes at 4 ^{0}C, the supernatant was carefully removed using a pasteur pipette and the tube was inverted over a paper towel to drain off residual liquid. Pellets were resuspended in 0.5 ml SM buffer and transferred to a 1.5 ml Eppendorf. RNase A was added to a final concentration 1 µg ml^{-1}, incubated at 37 ^{0}C for 20 minutes and centrifuged at 8,000 x g for 2 minutes at 4 ^{0}C. The supernatant was extracted with an equal volume of chloroform until the interface was clean (x3), 5 µl 10 % SDS and 5 µl 0.5 M EDTA (pH 8.0) were added to the supernatant and incubated at 65 ^{0}C for 15 minutes. The supernatant was extracted with phenol (x1), phenol / chloroform (x1) and chloroform (x1). Phage DNA was precipitated by addition of an equal volume isopropanol, incubated at -80 ^{0}C for 30 minutes, thawed and the phage DNA collected by centrifugation at 12,000 x g for 15 minutes at 4 ^{0}C, washed with 70 % ethanol, dried under vacuum and resuspended in 50 µl TE (pH 8.0). An aliquot (2 µl) of purified phage DNA was analysed by agarose gel electrophoresis to ascertain purity and concentration.

Intermediate scale preparation of phage lambda DNA using QIAGEN-tip 100 (from QIAGEN Inc., Chatsworth, CA 91311 USA).

QIAGEN-tips contain an anion exchange resin attached to silicagel which has been covalently coated with a hydrophilic substance to prevent nonspecific binding.

This protocol utilises an optimised buffer system for the "PEG-method" in combination with QIAGEN-tip 100. Plate lysates of the lambda were prepared as described by Sambrook *et al* (1989). Chloroform (1 µl ml^{-1}) was added to the plate lysate and the lysate was centrifuged to remove bacterial debris or insoluble agarose contaminants. The cleared supernatant was transferred to a 100 ml centrifuge tube (MSE), 100 µl of buffer L1 (20 mg ml^{-1} RNase A, 6 mg ml^{-1} DNase I, 0.2 mg ml^{-1} BSA, 10 mM EDTA, 100 mM Tris/HCl, 300 mM NaCl, pH 7.5) was added to the supernatant, mixed gently by inversion and incubated at 37 °C for 30 minutes. An aliquot (10 ml) of ice cold buffer L2 (30 % polyethylene glycol 6000, 3 M NaCl) was added, mixed gently by inversion and incubated on ice for 60 minutes. Phage particles were collected by centrifugation at 10,000 x g, the supernatant was discarded and the tube placed upside down on tissue paper for 1 minute to drain off residual liquid. The phage pellet was resuspended in 3 ml of buffer L3 (100 mM Tris/HCl, 100 mM NaCl, 25 mM EDTA, pH 7.5) using a pipetter, 3 ml of buffer L4 (4 % SDS) was added, mixed gently by inversion and incubated at 70 °C for 20 minutes then cooled on ice. An aliquot (3 ml) of buffer L5 (2.55 M potassium acetate, pH 4.8) was added, mixed gently by inversion and centifuged at 4 °C for 30 minutes (15,000 x g). The supernatant was transferred to 2 x 30 ml corex tubes and recentrifuged at 4 °C for 30 minutes (15,000 x g) and the particle-free lysate was transferred to a sterile 50 ml falcon tube. A QIAGEN-tip 100 was equilibrated with 3 ml buffer QBT (0.75 M NaCl, 50 mM MOPS, 15 % ethanol, pH 7.0 0.15 % Triton X-100), the cleared supernatant was loaded onto the QIAGEN-tip and allowed to enter the resin by gravity flow, washed with 10 ml buffer QC (1.0 M NaCl, 50 mM MOPS, 15 % ethanol, pH 7.0) and the DNA eluted with 5 ml buffer QF (1.25 M NaCl, 50 mM MOPS, 15 % ethanol, pH 8.2). The eluted DNA was transferred to a 15 ml corex tube, precipitated with 0.7 volume of isopropanol, incubated at room temperature for 15 minutes and centrifuged at room temperature for 30 minutes at 12,000 x g. The DNA pellet was washed with 10 ml 70 % ethanol, air dried for 5-10 minutes and resuspended in 100 µl sterile milli Q water. An aliquot (2 µl) analysed by agarose gel electrophoresis to ascertain the purity and concentration of the isolated phage DNA.

Agarose gel electrophoresis of DNA.

For most applications 0.7-1.2 % (w/v) agarose gels were used which are recommended for the separation of DNA fragments in the range of 10-0.25 kb (Sambrook *et al.* 1989). TAE buffer was used with gels and running buffer containing 1 µg ml^{-1} ethidium bromide in order to visualise the DNA bands (Sharp *et al* 1973). Samples were mixed with 0.5 volume of agarose gel loading buffer prior to loading. Commercially available restriction digests of lambda DNA were run as size markers. Agarose gels were electrophoresed for 4 hrs at 100 V or overnight at 30 V. DNA was visualised under long wave ultraviolet light (320 nm) and photographed.

Isolation of restriction fragments and PCR products from agarose gels.

Two methods were employed for isolating DNA from agarose gel.

Electroelution

Agarose gel slices containing DNA to be isolated were placed in prepared dialysis tubing using a minimum volume of buffer (Sambrook *et al.*. 1989) The tubing was placed in a minigel apparatus, perpendicular to the polarity and covered with TAE buffer and DNA was electrophoresed from the agarose gel for 20 minutes at 60 mA. After observation under uv, to check that the DNA had migrated to one edge of the tubing, the polarity was reversed for 20 seconds. Buffer containing the DNA was removed from the dialysis tubing, the DNA was extracted with phenol/chloroform, precipitated with 1 μl (20 μg μl^{-1}) glycogen carrier, 1/25 volumes 5 M ammonium acetate 0.25 M $MgCl_2$ pH 5.2 and three volumes of 100 % ethanol (-20 °C). The resulting DNA pellet was resuspended in 10 μl sterile milli Q water.

Silica Fines

Agarose gel slices containing DNA to be isolated were placed in Eppendorf tubes. A volume of sodium iodide (NaI) solution, which was three times the weight of the agarose gel slice, was added and the tube incubated at 65 °C for 10 minutes (until all the agarose had melted). The tube was cooled to room temperature and the DNA recovered by binding to 5 μl silica fines (equivalent to "Finebind", Amersham International plc) at room temperature for 10 minutes. Silica fines were pelleted by centrifugation at 12,000 x g for 15 seconds and washed by resuspension in 1 ml 70 % ethanol. Silica fines were repelleted and the wash solution aspirated and discarded. DNA was eluted by the addition of 50 μl TE buffer (pH 8.0) and incubation at 37 °C for 10 minutes. Silica fines were pelleted by centrifugation at 12,000 x g for 15 seconds and the eluted DNA transferred to a fresh Eppendorf tube for use in ligations.

Cloning procedures for isolated restriction fragments and PCR products

DNA fragments generated by restriction enzymes and isolated by electroelution were cloned into the vector pUC 18 (0.1-0.2 μg) restricted with enzymes to produce compatible termini for ligation. The restricted vector DNA was extracted with phenol/chloroform and precipitated with 1 μl (20 μg μl^{-1}) glycogen carrier, 1/25 volumes 5 M ammonium acetate 0.25 M $MgCl_2$ pH 5.2 and three volumes of 100 % ethanol (-20 °C). Ligations containing equimolar amount of DNA fragment and vector were prepared in a total volume of 10 μl containing 1 unit of T4 DNA ligase, a buffer supplied with the enzyme and incubated overnight at 15 °C.

PCR generated products were isolated by electroelution. Purified products were treated with T4 polymerase, to blunt the ends of the products, in the presence of a dNTP mix and a buffer supplied with the enzyme (Sambrook *et al.* 1989). The products were extracted with phenol/chloroform, precipitated with 1/10 volume sodium acetate pH 5.2 and three volumes of 100 % ethanol. The T4 polymerase treated products were cloned into the vector pUC 18 (0.1-0.2 μg) restricted with an enzyme to produce compatible blunt end termini. The restricted vector DNA was extracted with phenol/chloroform and precipitated with 1 μl (20 μg μl^{-1}) glycogen carrier , 1/25 volumes 5 M ammonium acetate 0.25 M $MgCl_2$ pH 5.2 and three volumes of 100 % ethanol (-20 °C). Ligations containing equimolar amount of DNA fragment and vector were prepared in a total volume of 10 μl containing 1 unit of T4 DNA ligase, a buffer supplied with the enzyme and incubated overnight at 15 °C.

As an alternative to treating PCR generated products with T4 polymerase for cloning purposes, a pGEM-T vector system supplied by Promega may be employed. The pGEM-T vector was restricted with *Eco* RV and 3' terminal thymidine was added to both ends. These single 3'-T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR generated product into a plasmid. More specifically, ligation of the overhangs takes advantage of the non-template dependent addition of a single deoxynucleotide, most commonly deoxyadenosine, to the 3'-end of PCR products by certain thermostable polymerases. PCR generated products were isolated from agarose gel by a method described by Zhen and Swank (1993), (for further details refer to preparation of PCR products for sequencing. DNA was extracted with phenol/chloroform and precipitated with 1/10 volume 3 M sodium acetate (pH 5.2) and 3 volumes of 100 % ethanol (-20 °C). Washed (x2) with 70 % ethanol and the pellet resuspended in 10 μl of sterile milli Q water. Ligations containing equimolar amount of DNA fragment and vector were prepared in a total volume of 10 μl containing 1 unit of T4 DNA ligase, a buffer supplied with the enzyme and incubated overnight at 15 °C.

Plasmid transformation of competent cells.

Commercially available DH5α competent cells were transformed using the following procedure: an aliquot of ligated plasmid (2-10 ng was added to freshly thawed subcloning efficiency competent cells, mixed gently and incubated on ice for 20 minutes. Cells were heat shocked at 42 °C for 2 minutes, four volumes of SOC medium was added and the cells incubated shaking at 37 °C for 1 hour to express ampicillin resistance. Aliquots were spread onto LB agar plates containing X-gal and ampicillin, inverted and incubated at 37 °C overnight.

For use with the pGEM-T vector system JM101 or JM109 competent cells were transformed using the following procedure: an aliquot of ligated plasmid (2-10 ng was added to freshly thawed high efficiency competent cells, mixed gently and incubated on ice for 20 minutes. Cells were heat shocked at 42 °C for 50 seconds, 1.4 ml LB medium was added and the cells incubated shaking at 37 °C for 1 hour to express carbenicillin resistance. Aliquots were spread onto LB agar plates containing X-gal, IPTG and carbenicillin, inverted and incubated at 37 °C overnight.

Size fractionation of genomic DNA.

Genomic DNA (100 μg) was digested with a restriction endonuclease for 6 hours (the contents of the tube were mixed every 30 minutes). The digest was then extracted with phenol/chloroform, precipitated with 1/10 volume sodium acetate pH 5.2 and three volumes of 100 % ethanol and resuspended in sterile milli Q water. The resuspended DNA was heated at 68 °C for 10 minutes before loading on a sucrose gradient.

A 30-60 % sucrose density gradient was produced in Beckman polyallomer tubes. Sucrose solutions were prepared in a buffer containing 1 M NaCl, 20 mM Tris-Cl, pH 8.0, 5 mM EDTA. The DNA sample was fractionated by sucrose density gradient centrifugation at 20,000 x g for 18 hours at 20 °C in a MSE prepspin centrifuge, using a 43127-104 MSE swing out rotor. The centrifuge tube was then punctured at the bottom and 1 ml fractions collected. Aliquots (200 μl) of each fraction was loaded onto a dot-blot and hybridised to a radiolabelled probe.

Titration and screening of a cosmid library.

A host bacterial strain was grown overnight at 37 °C in LB (10 ml) containing 0.2 % maltose and 10 mM $MgCl_2$. Cells were collected by centrifugation at 4,000 x g, resuspended in 0.4 volume 10 mM $MgCl_2$ and stored at 4 °C until required. An aliquot (2 μl) of the cosmid library was diluted to 100 μl with SM buffer, an equal volume of the previously prepared host bacterial strain was added and incubated at 37 °C for 20 minutes. LB (4.8 ml) was added and the cells incubated at 37 °C for 45 minutes to express tetracycline resistance. Aliquots (0 - 100 μl) of the cells were spread over LB plates containing tetracycline, incubated overnight at 37 °C and colonies counted. Using the results from the titration approximately 3 x 10^4 cfu were screened. Large (200 mm x 200 mm) plates were prepared by using 2 μl of library as previously described, a total of 1.65 ml of cells was spread onto 3 large plates, prewarmed to 37 °C. The plates were incubated at 37 °C overnight then cooled to 4 °C. Nitrocellulose filters were placed onto the surface of the top agar, and using a hole punch and ink, the filters were orientated. The filters were processed (refer to lysis of bacterial colonies on nitrocellulose filters for details), hybridised with radiolabelled probe and exposed to X-ray film. Colonies which showed relatively strong hybridisation to the probe, when the signals on the X-ray film were aligned with the original agar plates, were streaked onto a reference plate, incubated overnight at 37 °C then stored at 4 °C. Due to the colony density on the original plate it was difficult to remove a single colony without taking other weakly hybridising colonies as well. Secondary and tertiary screens were performed so that individual hybridising colonies could be isolated. A colony screen described by Sambrook *et al* (1989) was performed on all colonies streaked onto the reference plate. One filter was processed (refer to lysis of bacterial colonies on nitrocellulose filters for details), hybridised to a radiolabelled probe and exposed to X-ray film.

Titration and screening of lambda phage libraries.

It is important to determine the phage titre of the library to be screened in order to screen a representative amount of plaques from the library. The optimal plaque forming units (pfu) required for screening is typically 2-3 fold higher than that required to produce a dense field of individual plaques without total lysis of the bacterial lawn. A host bacterial strain was grown overnight at 37 ^{0}C in LB (10 ml) containing 0.2 % maltose and 10 mM $MgCl_2$. Cells were collected by centrifugation at 4,000 x g, resuspended in 0.4 volume 10 mM $MgCl_2$ and stored at 4 ^{0}C until required. An aliquot (10 μl) of phage suspension was diluted with SM buffer (1:10) and 50 μl of these dilutions were absorbed to an equal volume of a host bacterial strain at 37 ^{0}C for 20 minutes. Cells were mixed with 3 ml of LB top agarose at 50 ^{0}C, poured over prewarmed (37 ^{0}C) LB agar, incubated overnight at 37 ^{0}C and the plaques counted.

Approximately 1.5 x 10^4 pfu were incubated with 800 μl host cells for 20 minutes at 37 ^{0}C, then added to 45 ml of 0.7 % LB top agarose, poured over 2 day old large (200 mm x 200 mm) LB plates, prewarmed to 37 ^{0}C and the plates were incubated at 37 ^{0}C overnight then cooled to 4 ^{0}C. Nitrocellulose filters (in duplicate) were placed onto the surface of the top agar, and using a hole punch and ink, the filters were orientated. The filters were processed (refer to lysis of phage plaques on nitrocellulose filters for details), hybridised with radiolabelled probe and exposed to X-ray film. Plaques which showed duplicate hybridisation to the probe, when the signals on the X-ray film were aligned with the original agar plates, were removed as plugs with the wide end of a pasteur pipette and stored in 1 ml SM buffer at 4 °C. Due to the plaque density on the original plate it was never possible to remove a single plaque without taking other weakly hybridising plaques as well. Secondary and tertiary screens were performed so that an individual hybridising plaque could be isolated. The SM buffer containing the agar plug was incubated at room temperature for at least 2 hours to allow diffusion of the phage particles. Dilutions of the phage suspension were prepared as previously described, the LB top agarose containing diluted phage and host cells were poured over small LB plates. Plates which had well seperated plaques were chosen for transfer to nitrocellulose filters. These filters were screened, single hybridising plaques were removed using the thin end of a pasteur pipette, transferred to 1 ml SM buffer containing 20 µl of chloroform, to prevent bacterial growth, and stored at 4 ^{0}C.

Lysis of bacterial colonies on nitrocellulose filters.

The method used was described by Sambrook *et al.* (1989), either for screening genomic libraries (as previously described) or for screening large numbers of putative transformants.

Nitrocellulose filters (marked with a grid) were placed on top of LB agar plates containing the appropriate antibiotic selection. Putative transformants were transferred, in duplicate, to the filters using sterile cocktail sticks and the plates incubated overnight at 37 ^{0}C. The duplicate plates were marked in six identical positions in order to identify hybridising colonies. One of the duplicate nitrocellulose filters or the filter from the library screen were removed and air dried on Whatman 3 MM paper. The cells were lysed by placing the filters in turn, each for 5 minutes, onto 3 MM paper soaked in the following solutions, drying on 3 MM paper between each step; 0.5 M NaOH (x2); 1 M Tris-HCl pH 7.4 (x1); 1.5 M NaCl, 0.5 M Tris-HCl pH7.4 (x1). The filters were air dried for 30 minutes and then baked in a vacuum oven at 80 ^{0}C for 2 hours.

Lysis of phage plaques on nitrocellulose filters.

The method used was described by Sambrook *et al.* (1989) for screening genomic libraries.

Nitrocellulose filters (in duplicate) were placed onto the surface of the top agar, and using a hole punch and ink, the filters were orientated. The nitrocellulose filters were removed, air dried on Whatmann 3 MM paper and the phage were lysed by placing the filters in turn, each for 5 minutes, onto 3 MM paper soaked in the following solutions, drying on 3 MM paper between each step; 1.5 M NaCl, 0.5 M NaOH; 1.5 M NaCl, 0.5 M Tris-HCl pH 8.0; 3 x SSC. The filters were air dried for 30 minutes and then baked in a vacuum oven at 80 ^{0}C for 2 hours.

Transfer of DNA to nitrocellulose membranes

The method was a modification of that of Southern (1975).

Gels containing nucleic acids were blotted onto nitrocellulose filters as described by Sambrook *et al.* (1989). Gels were soaked in denaturating solution for 2 x 30 minutes (1.5 M NaCl, 0.5 M NaOH) followed by 2 x 30 minutes in neutralising solution (1.5 M NaCl, 0.5 M Tris.Cl pH 7.0, 1 mM EDTA). The gel was blotted onto nitrocellulose overnight as described by Sambrook *et al.* (1989) except that nappy liners were used instead of paper towels and 20 x SSC was used as the transfer buffer. After blotting, the position of the wells were marked on the filter with ink, DNA was bound to the filter by baking for 2 hours at 80 ^{0}C under vacuum and stored at room temperature until required.

Random primed labelling of DNA.

An Amersham megaprime kit was used for these reactions. Approximately 300 ng DNA (prepared by either restriction or PCR and electroeluted) in a total volume of 28 μl was boiled for 5 minutes, transferred to ice, 10 μl of labelling buffer, 5 μl of primers, 5 μl ^{32}P-dCTP and 2 μl of Klenow enzyme were added and incubated at room temperature for 2-4 hours.

The labelling reaction was stopped by the addition of 5 μl of Stop-Dye (60 mM EDTA, 100 mg ml^{-1} blue dextran, 1 mg ml^{-1} xylene cyanol). The unincorporated label was removed using Bio-Rad Biospin 30 columns. The column was inverted several times to resuspend the resin, the excess buffer was drained by gravity and discarded. The column and collection tube were centrifuged for 2 minutes in a swinging bucket rotor centrifuge at 1,500 x g. The reaction was loaded to the centre of the resin, centrifuged for 4 minutes and the radiolabelled DNA collected in a fresh collection tube. The unincorporated label remains on the resin. An aliquot (1 μl) was counted with Ecoscint scintillant. The radiolabelled DNA was boiled for 5 minutes before use to render them single-stranded.

Hybridisation of radiolabelled DNA probes to filter-immobilised DNA.

All prehybridisation and hybridisation reactions were performed in hybridisation tubes and ovens supplied by Techne (Cambridge) Ltd. Southern blots and *in situ* filters were incubated at 65 °C for 2-4 hours in prehybridisation solution (6 x SSC, 1 x Denhardts, 0.5 % SDS, 0.05 % PPi, 0.1 % denatured herring sperm DNA) This solution was replaced with hybridisation solution at 65 ^{0}C (6 x SSC, 1 x Denhardts, 0.5 % SDS, 0.05 % PPi, 1 mM EDTA), denatured labelled probe was added and the hybridisation reaction performed overnight at 65 ^{0}C. Following hybridisation the filters were washed sequentially for 20 minutes at 65 ^{0}C with the following solutions until the required stringency was reached: 2 x SSC 0.1 % SDS (x2), 1 x SSC 0.1 % SDS (x2), 0.1 x SSC 0.1 % SDS (x2). Filters were sealed in polythene bags and exposed to x-ray film.

Autoradiography.

Nitrocellulose filters were sealed in polythene bags and the following procedure carried out in a dark room under a red safe-light: X-ray film was preflashed, placed between the filter and an intensifying screen within a cassette and left to be exposed. For periods longer than 6 hour the cassette was stored at -80 °C. X-ray films were developed in Ilford phenisol developer for a maximum 5 minutes, rinsed in water and immersed in Kodak unifixer for 3 minutes, rinsed in running water for 10 minutes and air dried.

Determination of DNA concentration.

DNA concentration was determined on a Pye Unicam SP-800 dual beam spectrophotometer. For single-stranded DNA, a 1 mg ml^{-1} solution was assumed to give an OD_{260} of 33 and for double-stranded DNA, a 1 mg ml^{-1} solution was assumed to give an OD_{260} of 20. Alternatively, the concentration and purity of a DNA sample was determined by agarose gel electrophoresis with a standard of known concentration (pGEM 3Z).

PROCEDURES USED WITH POLYMERASE CHAIN REACTIONS.

Unless otherwise stated all general procedures were performed as described by Sambrook *et al.* (1989).

Oligonucleotide synthesis.

Oligonucleotides were synthesised using an Applied Biosystems 381A DNA Synthesiser operated with a standard protocol detailed in the user manual. Oligonucleotides required for DNA sequencing were ethanol precipitated prior to quantitation (refer to optimisation of DNA sequencing for further details). Oligonucleotides required for polymerase chain reactions were quantified and diluted to a suitable working solution (20 μM). Oligonucleotides required for DNA sequencing were quantified and diluted to a suitable working solution (double-stranded templates 3.2 pMol μl^{-1}).

Oligonucleotides designed for use with the polymerase chain reaction.

Pisum sativum (Pea).

Three sets of oligonucleotides were designed based upon the genomic sequence of pλS2 (designated $PsMT_A$) isolated from *Pisum sativum* . These oligonucleotides were designed to amplify other members of a small multi gene family (Evans *et al.* 1990).

Primer set 1

Primer 190 5' GGC GAA TTC TGC AAI TGT GGI GAI AIC TGC A 3'

Primer 191 5' GGC GAA TTC GCA GTT GCA IGG ITC ACA IIT GCA GTT 3'

These two primers were designed from DNA sequences obtained from *P. sativum* (Evans *et al.* , 1990) and *Mimulus guttatus* (De Miranda *et al* ., 1990) where there was a base change on the third nucleotide of the amino acid codon. A three base extension (GGC) and an *Eco* RI restriction site (GAATTC) were added to the 5' end of both primers.

Primer set 2

Primer 238 5' GGC GAA TTC AAI TG(TC) GGI GA(TC) (TA)I(TC) TG(TC) AA 3'

Primer 239 5' GGC GAA TTC GTI CAI TT(AG) TCI CC(AG) CA(TC) TT

These two primers were designed based on the amino acid sequence predicted from the *P. sativum* genomic sequence ($PsMT_A$). The base analogue, inosine, was incorporated when more then two nucleotides represented a particular amino acid, giving rise to a high degree of redundancy. A three base extension (GGC) and an *Eco* RI restriction site (GAATTC) were added to the 5' end of both primers.

Primer set 3

Primer 283 5' GGC GTC GAC TTT GCA GTT GCA AGG GTC AC 3'

Primer 284 5' GGC GAA TTC GTC TGG ATG TGG TTG TTG TGG 3'

These two primers were perfectly matched to the published sequence of the *P. sativum* genomic sequence ($PsMT_A$). A three base extension (GGC) and the restriction sites *Sal* I (GTCGAC) and *Eco* RI (GAATTC) were added to the 5' end of the primers respectively

Primers for anchored polymerase chain reaction (APCR).

Primer 301 5' TCT CCA AAG AAT CAT CTA TAT GAT TA 3'

This primer was designed to the intron of the *P. sativum* genomic sequence $PsMT_A$.

Primer 302 5' CCA ACA GAT TAA TAA CCA TAT AG 3'

This primer was designed to the intron of *PsMT*$_B$ amplified by PCR from *P. sativum* genomic DNA.

***Arabidopsis thaliana* (Thale cress).**

These two primers were designed to the cDNA sequence characterised by Takahashi. K. (1991), obtained from the EMBL database (accession number X628818)

Primer 518 5' GGC GAA TTC ACT TGC AGG TGC AAG GAT C 3'

A three base extension (GGC) and an *Eco* RI restriction site (GAATTC) was added to the 5' end of this primer.

Primer 519 5' CGT GGG ATC CCC ATG TCT TGC TGT GGA GGA 3'

A four base extension (CGTG) and a *Bam* HI restriction site (GAATTC) was added to the 5' end of this primer.

***Arabidopsis thaliana*.**

These two primers were designed to the cDNA sequence characterised by Zhou and Goldsbrough, (1994), obtained from the GENBANK database (accession number L15389).

Primer 826 5' CGG CAC GAG GAA GAA ACT AC 3'

Primer 827 5' CGG TCA CGT CTT TAT TCC AAC 3'

***Arabidopsis thaliana*.**

These two primers were designed to the cDNA sequence characterised by Raynal *et al.*, (1993), obtained from the GENBANK database (accession number Z27049).

Primer 900 5' GGC AAA GGA AGT GCA AGC GCT 3'

Primer 901 5' GAT CGA CTA GCT AGT AGT ACC 3'

Amplification of cDNA and genomic DNA using the polymerase chain reaction

PCR reactions were performed essentially as described by Saiki *et al.* (1988). To avoid contamination by foreign template DNA's all reaction tubes and tips were autoclaved and a positive displacement pipetteman from Gilson Ltd. was used to prepare the reactions. Standard 100 µl reactions and general cycling parameters were used for most of the reactions, any alterations have been specified. The reagents were mixed on ice in 0.5 ml Eppendorf tubes.

Standard reaction.

X µl sterile Milli Q water, 10 µl 10 x reaction buffer, 16 µl stock dNTPs (1.25 mM of each), 5 µl 5' Primer (20 µM), 5 µl 3' Primer (20 µM), 1 µl *Taq* polymerase (5 Units/µl) and template DNA (20 ng - 1.5 µg). The reaction was mixed thoroughly, centrifuged briefly and overlaid with 100 µl mineral oil.

Commercial dNTP's (100 mM), 12.5 µl of each dNTP stock were added to 950 µl of sterile dH_2O and mixed thoroughly. The amount of template required for PCR varied depending on the size and source of the DNA. For instance, when amplifying a segment of DNA within a plasmid, 20 ng of template was sufficient due to the small amount of non target DNA present. However, when amplifying a segment of DNA present in total genomic DNA, 1.0 - 1.5 µg of template was required due to the large amount of non target DNA present. When using cDNA and genomic libraries as templates 5 µl of the library was used in a standard PCR reaction. A negative control reaction (no tempate DNA), and where possible, positive control reactions were prepared.

Cycling parameters.

All polymerase chain reactions were performed using a Pharmacia.LKB Gene ATAQ Controller.

Plasmids:	9 cycles	94 °C for 1.5 min., 45 °C for 1 min., 72 °C for 2 min.
	1 cycle	94 °C for 1.5 min., 45 °C for 1 min., 72 °C for 5 min.
Genomic:	29 cycles	94 °C for 1.5 min., 45 °C for 1 min., 72 °C for 2 min.
	1 cycle	94 °C for 1.5 min., 45 °C for 1 min., 72 °C for 5 min.
Library:	32 cycles	94 °C for 1.5 min., 45 °C for 2 min., 72 °C for 2 min.
	1 cycle	94 °C for 1.5 min., 45 °C for 2 min., 72 °C for 5 min.

Once thermal cycling was completed, 20 µl of the reaction products were separated by agarose gel electrophoresis and the remainder of the reaction stored at -20 °C.

OPTIMISED PROCEDURES FOR AUTOMATED DNA SEQUENCE ANALYSIS.

A number of modifications have been made to the procedures detailed in the Applied Biosystems 373A DNA Sequencer User Manual and are outlined below.

Automated DNA sequence analysis was performed on an Applied Biosystems 373A DNA Sequencer, using a modification of the dideoxy chain termination method. All sequencing reactions were performed on either a Pharmacia LKB Gene ATAQ Controller (dye-deoxy terminator chemistry) or a Techne PHC-3 (96-well) thermal cycler (dye primer chemistry). The cycling parameters have been optimised for use with these two instruments, which differ from those detailed in the Applied Biosystems user manual (refer to optimisation of DNA sequencing for details). DNA sequencing reactions were performed on double-stranded DNA templates (plasmid DNA, cosmid DNA and PCR generated DNA). A number of protocols for the preparation of DNA templates for automated sequence analysis have also been optimised (refer to optimisation of DNA sequencing for details). The sequencing reactions were separated by 6 % poly-acrylamide gel electrophoresis.

Template preparation.

It has long been established that the quality of the sequence data obtained is directly proportional to the quality and quantity of the DNA template supplied. A number of template preparations have been under investigation to identify suitable methods for automated DNA sequence analysis. Three methods for obtaining plasmid DNA for sequencing are detailed below and must be followed exactly to ensure successful sequencing.

Modified Birnboim and Doly (1979) alkaline lysis with polyethylene glycol (PEG) precipitation.

All stock solutions were prepared in milli Q water using analytical grade chemicals and autoclaved.

A 10 ml culture of the bacteria was grown overnight at 37 °C in LB containing the appropriate antibiotic selection. All the cells were harvested by centrifugation for 5 minutes at 1,500 x g in a MSE bench top centrifuge (swing out rotor). The supernatant was removed and the pellet partially dried by inverting the tube over tissue paper, 100 µl of ice cold solution 1 (50 mM Glucose, 10 mM EDTA, 25 mM Tris-Cl, pH 8.0) was added, the cells were resuspended using a pipette and incubated at room temperature for 5 minutes. An aliquot (200 µl) of freshly prepared solution 2 (0.2 N NaOH, 1.0 % SDS) was added, mixed gently by inversion and incubated on ice for 5 minutes. An aliquot (150 µl) of ice cold solution 3 (11.5 ml glacial acetic acid, 28.5 ml water, 60 ml 5 M potassium acetate) was added, mixed by inversion and incubated on ice for 5 minutes. Chromosomal DNA and other cellular debris were pelleted by centrifugation at 12,000 x g for 3 minutes, transferred to a fresh tube and recentrifuged at 12,000 x g for 3 minutes. The supernatant was transferred to a fresh tube, RNase A was added to a final concentration of 20 µg μl^{-1} and incubated at 37 °C for 20 minutes. An equal volume of phenol/chloroform (1:1) was added to the RNase A treated supernatant, the tube was vortexed for 30 seconds and centrifuged for 1 minute at 12,000 x g. The aqueous phase was transferred to a fresh tube, 3 volumes of 100 % ethanol was added and the DNA precipitated at -20 °C for 2 hours or -80 °C for 20-30 minutes. The DNA was pelleted by centrifugation at 12,000 x g. The pellet was washed (x2) with 200 µl 70 % ethanol, air dried and resuspended in 16 µl sterile milli Q water. 2 M NaCl (8 µl) was added, mixed, then 13 % PEG (20 µl) was added, mixed thoroughly and incubated on ice for 20-30 minutes. The DNA precipitate was collected by centrifugation at 12,000 x g for 15 minutes, the pellet was washed (x3) with 200 µl 70 % ethanol, air dried and resuspended in 20 µl sterile milli Q water. The purity and concentration of the DNA sample was determined by agarose gel electrophoresis using a known standard (pGEM 3Z).

Promega Magic / Wizard Minipreps (improved protocol).

A 10 ml culture of the bacteria was grown overnight at 37 °C in LB containing the appropriate antibiotic selection. All the cells were harvested by centrifugation for 5 minutes at 1,500 x g in a MSE bench top centrifuge (swing out rotor). The supernatant was removed and the pellet partially dried by inverting the tube over tissue paper, 300 µl of ice cold resuspension buffer (50 mM Tris, pH 7.5 10 mM EDTA, 100 µg/ml RNase A) was added, the cells were resuspended using a pipette and the cells transferred to a 1.5 ml microcentrifuge tube. An aliquot (300 µl) of cell lysis buffer (0.2 N NaOH, 1.0 % SDS) was added and mixed gently by inversion until the cell suspension cleared. An aliquot (300 µl) of neutralisation buffer (1.32 M potassium acetate) was added and mixed by inversion. Chromosomal DNA and other cellular debris were pelleted by centrifugation at 12,000 x g for 3 minutes, transferred to a fresh tube and recentrifuged at 12,000 x g for 3 minutes. The supernatant was transferred to two microcentrifuge tubes (450 µl/tube), 500 µl Magic/Wizard Miniprep DNA Purification Resin was added, mixed by inversion and incubated at room temperature for 5 minutes. A Magic / Wizard Minicolumn was prepared as described by the manufacturer's instructions. The resin/DNA mix from both tubes was added to the syringe barrel, the syringe plunger was inserted and the slurry was gently pushed into the Minicolumn. The syringe was detached from the Minicolumn and the plunger removed from the syringe barrel. The syringe barrel was reattached to the Minicolumn and the resin washed with 3 ml Wash Solution (0.2 M NaCl, 20 mM Tris, pH 7.5, 5 mM EDTA and an equal volume of 95 % ethanol was added before use). The Minicolumn was removed from the syringe barrel, transferred to a 1.5 ml microcentrifuge tube and centrifuged at 12,000 x g for 1 minute. The Minicolumn was transferred to a fresh collection tube, 100 µl of sterile milli Q water heated to 65-70 °C was added and incubated for 1 minute at room temperature. The DNA was eluted by centrifugation at 12,000 x g for 1 minute. The purity and concentration of the DNA sample was determined by agarose gel electrophoresis using a known standard (pGEM 3Z).

Qiagen QIAprep-spin Plasmid Kit.

There was only one modification made to this protocol. When the DNA was eluted from the QIAprep-spin column, 100 µl sterile milli Q water at 55 °C was used in place of TE or Tris buffer at room temperature. The purity and concentration of the DNA sample was determined by agarose gel electrophoresis using a known standard (pGEM 3Z).

Purification of PCR products for DNA sequencing.

It is very important to obtain a homogeneous PCR product, free of any primers, dNTPs and any contaminating non-specific PCR products for sequencing. A number of methods have been evaluated and three methods for purifying PCR products for sequencing are currently in use.

Purification of PCR products from agarose gels (Zhen and Swank, 1993).

This method developed by Zhen and Swank (1993) is a modification of the procedure developed by Hogness described in Sambrook *et al.* (1989) for the purification of size-fractionated DNA fragments directly from agarose gels. The fragments or PCR generated products are purified in the presence of polyethylene glycol (PEG).

A PCR product was monitored as it migrated through an agarose gel containing ethidium bromide using a long-wave uv light. After sufficient resolution, the product was allowed to migrate into a buffer-filled trough (300 µl - 450 µl of a stock solution of: 15 % PEG-8000, 2 x TAE, ethidium bromide 0.5 µg ml^{-1}) cut using a sharp scalpel blade, directly in front of its path of migration. Electrophoresis was continued (25 V) and the mobility of the product was checked periodically with long-wave uv light. Once the product had moved into the centre of the trough, electrophoresis was stopped and the DNA-containing PEG-TAE solution was pipetted into a microcentrifuge tube. The PCR product was purified by extracting (x2) with an equal volume of phenol/ chloroform, the aqueous phase was precipitated with 1/10 volume of 3 M sodium acetate (pH 5.2), 2 volumes of 100 % ethanol and incubated at -20 °C for 2 hours or -70 °C for 20-30 minutes. The precipitate was pelleted by centrifugation at 12,000 x g for 15 minutes, washed (x3) with 200 µl of 70 % ethanol. The DNA pellet was resuspended in 10 µl sterile milli Q water overnight at 4 °C. The purity and concentration of the DNA sample (1 µl) was determined by agarose gel electrophoresis using a known standard (pGEM 3Z).

Purification of PCR products from agarose gels using DEAE cellulose.

The procedure for the transfer of DNA fragments to DEAE-cellulose has been described by Sambrook et al. (1989) and has been modified for the purification of PCR products for sequencing (J.V. Hookey personal communication).

On ice water, 100 μl chloroform and 100 μl phenol (100 g phenol, 25 ml 2 M Tris-Cl, pH 8.0, 32 ml milli Q water, 6 ml m-cresol and 125 mg 8-hydroxyquinoline) were added to 100 μl of PCR product in an Eppendorf, mixed thoroughly using a vortex for 30 seconds and centifuged at 12,000 x g for 1 minute. The upper aqueous phase was removed, 0.1 volume of 3 M sodium acetate, pH 5.3, 3 volumes of cold 100 % ethanol was added, mixed by inversion and incubated overnight at -20 °C. The precipitate was pelleted by centrifugation at 12,000 x g for 15 minutes, washed (x2) with 70 % ethanol, air dried and resuspended in 100 μl sterile milli Q water. All of the PCR product was loaded onto a 1 % agarose gel and electrophoresed for 1-2 hours. The product was visualised by long wave uv light and an incision made with a scalpel blade directly in front of the leading edge of the fragment to be purified. DEAE-cellulose membranes (Scheicher and Schuell, NA-45) were cut into squares (approx. 5 mm x 5 mm), soaked for 5 minutes in 10 mM EDTA, pH8.0, transferred to 0.5 N NaOH and soaked for a further 5 minutes. The membranes were washed (x6) with sterile milli Q water and stored for up to 2 weeks at 4 °C. A prepared membrane was then inserted into the incision previously made in the gel. Electrophoresis was resumed at 150 mA for 5 - 10 minutes and the migration of the product checked with a uv light. The membrane was removed and rinsed in 10 ml of low-salt buffer (50 mM Tris-Cl, pH 8.0, 0.15 M NaCl, 10 mM EDTA, pH 8.0) for ten minutes (it is important not to let the membranes dry as the DNA will bind irreversibly). The membrane was transferred to a microcentrifuge tube containing 400 μl of high-salt buffer (50 mM Tris-Cl, pH 8.0, 1 M NaCl, 10 mM EDTA, pH 8.0) and incubated at 65 °C for 30 minutes. The fluid was transferred to a fresh tube, a second aliquot of high-salt buffer was added to the membrane and incubated at 65 °C for a further 15 minutes. The aliquots were combined, extracted with an equal volume of phenol (x1) and an equal volume of chloroform (x1). The aqueous phase was transferred to a fresh tube, 0.2 volumes of 10 M ammonium acetate and 2 volumes of cold 100 % ethanol was added. The DNA was precipitated overnight at -20

°C. The precipitate was pelleted by centrifugation at 12,000 x g for 15 minutes, washed (x2) with 70 % ethanol, air dried and resuspended in 10 µl sterile milli Q water. The purity and concentration of the DNA sample (1 µl) was determined by agarose gel electrophoresis using a known standard (pGEM 3Z).

Commercially available PCR "clean-up" kits.

For relatively fast procedures for obtaining purified PCR products commercially available kits have been used. These include Promega Magic™ PCR Preps DNA Purification System for Rapid Purification of DNA Fragments and QIAGEN QIAquick-spin PCR Purification Kit. Essentially, the PCR fragments are bound to a resin and loaded onto a minicolumn (as detailed by the manufacturer). Washed to remove any primers and unincorporated nuleotides and the DNA eluted in sterile millQ water.

<u>DNA concentration for sequencing.</u>

The DNA concentration is also very important. A standard of known concentration should be loaded onto the agarose gel when quantifying samples for sequencing. Specific amounts of DNA in specific volumes are required depending on the sequencing chemistry used. When using cycle sequencing, the amount of DNA required is considerably less than for conventional sequencing.

Taq **Dye Primer Chemistry (universal primers).**

Total amount of double-stranded DNA required was 1.2 µg - 1.5 µg/6 µl and was supplied at a concentration of 200-250 ng μl^{-1}.

Taq **Dye-Deoxy™ Terminator Chemistry (custom primers).**

Total amount of double-stranded DNA required was 1 µg/10 µl and was supplied at a concentration of 100 - 200 ng μl^{-1}.

When sequencing directly from purified PCR products, the amount of DNA required was a molar ratio to the length of the product.

Custom Primers.

The design, purification and quantification of a primer to be used in automated DNA sequencing is very important. In recent months, careful examination of the performance of numerous custom primers has resulted in specific criteria which should be adhered to when designing custom primers:

Primer length	18-22 mer depending on GC content.
	(preferably 50 - 60 % GC rich).
If possible	G / C at both ends.
Exhibit	nearest-neighbour interactions *
	(dinucleotides within the primer: CG; GC; GG; CC).
	(De Bellis *et al.* 1992).
Avoid	runs of single bases especially at 3' end.
	(GGG at the 3' end appears to be very detrimental to the performance of the custom primer).

*For example: The M13 -21 universal primer (18 mer, 50 % GC)

5' TGTAAAA<u>CG</u>A<u>CGGCC</u>AGT 3'

This primer has five dinucleotides within the sequence of the primer CG; GC; GG; CC (one pair CG, then a block of 5 bases CGGCC which constitutes 4 individual dinucleotides CG; GG; GC; CC). The presence of these dinucleotides increases the stability of the primer, rather than 50 % GC and the bases occurring as mononucleotides. It is advisable to design a sequencing primer with 4 - 8 dinucleotides which is dependant on the actual primer length.

Purification of primers for sequencing.

Oligonucleotides were synthesised on an Applied Biosystems 381A DNA Synthesiser using 40 nM columns. After deprotection (detailed in the instruments user manual) the oligonucleotide in ammonia was transferred to a 2 ml screw cap Wheaton vial and incubated overnight at 55 °C. The vial was transferred to a fridge to cool, then the oligonucleotide in ammonia was divided into two 1.5 ml Eppendorf tubes and dried down in a vacuum centrifuge. Both aliquots of the oligonucleotide were resuspended in 360 µl of ammonia. 40 µl 3 M sodium acetate and 1.2 ml 100 % ethanol were added and incubated at -20 °C for 1 hour. The precipitated oligonecleotide was collected by centrifugation at 12,000 x g for 15 minutes, the pellets were washed (x2) with 1 ml 80 % ethanol, dried briefly under vacuum. One half of the oligonucleotide was stored at -20 °C as a quality control measure if required, the other half was resuspended in 100 µl sterile milli Q water. 5 µl was analysed by a spectrophotometer, the concentration of the oligonucleotide was calculated and 1 µl was diluted to a concentration suitable for double-stranded *Taq* dye-deoxy terminator DNA sequencing (3.2 pMol μl^{-1})

Sequencing chemistries and cycling parameters required for cycle sequencing.

All *Taq* Dye Primer Cycle Sequencing reactions were performed on a Techne PHC-3 96-well Thermal Cycler using the reagents and volumes specified in the Applied Biosystems DNA sequencing protocol (part no. 901482 Rev. B). This instrument was fitted with a heated lid and heat-sealing film was used to seal the tubes so as to eliminate the use of mineral oil to stop evaporation. However, this was not the case when using very small reaction volumes (5 and 10 µl) the oil was still required to stop evaporation.

All *Taq* Dye-Deoxy™ Terminator Cycle Sequencing reactions were performed on a Pharmacia LKB Gene ATAQ Controller 40-well Thermal Cycler using the reagents and volumes specified in the Applied Biosystems DNA sequencing protocol (part no. 901497 Rev. C).

However, the actual cycling parameters had to be optimised for these particular instruments because the cycling parameters detailed in the protocol were those optimised for Perkin-Elmer Cetus instruments. To optimise the instruments a time course for the annealing step was used to identify the parameters that gave the greatest signal intensity.

Techne PHC-3 96-well Thermal Cycler.

Rapid thermal ramp to 95 °C

95 °C 30 seconds

Rapid thermal ramp to 55 °C

55 °C 50 seconds

Rapid thermal ramp to 70 °C

70 °C 70 seconds

15 cycles total

Continue thermal cycling as follows:

Rapid thermal ramp to 95 °C

95 °C 30 seconds

Rapid thermal ramp to 70 °C

70 °C 70 seconds

15 cycles total

Rapid thermal ramp to 4 °C and hold.

Pharmacia LKB Gene ATAQ Controller 40-well Thermal Cycler.

Rapid thermal ramp to 96 °C

96 °C 30 seconds

Rapid thermal ramp to 50 °C

50 °C 50 seconds

Rapid thermal ramp to 60 °C

60 °C 4 minutes

25 cycles total

Rapid thermal ramp to 4 °C and hold.

When using these two instruments it should be noted that prolonged refrigeration reduces the life-span of the peltier systems. It is therefore, good practice to proceed to the next step within one hour of the cycling being completed.

Purification of sequencing reactions.

Dye Primer sequencing reactions.

To utilise this one lane four dye DNA sequencing system, the reactions (A, C, G and T) were combined in a single tube containing 100 µl 95 % ethanol and 2 µl 3 M sodium acetate, pH 5.2 and incubated on ice for at least 10 minutes (or may be stored at -20°C overnight). The precipitated extension products were collected by centrifugation at 12,000 x g, washed with 200 µl 70 % ethanol (for an experienced user this step may be omitted but extreme care must be taken to ensure removal of all the ethanol/sodium acetate mix. If any salt remains in the sample the first 50 bases in the sequencing ladder appear very narrow, giving rise to problems with automatic tracking in the analysis software). The pelleted extension products were air dried (up to 10 minutes), resuspended in 5 µl 50 mM EDTA : deionised formamide (1 : 5) and 4 µl / 6 µl (36-well / 24-well comb respectively) loaded onto a pre electrophoresed (10 minutes) 6 % acrylamide/ urea sequencing gel as described in the user manual.

Dye-Deoxy™ terminator sequencing reactions.

It is very important to remove all the unincorporated dye terminators. Two methods were employed, Bio-Rad Biospin 30 columns or phenol/chloroform extraction.

On a routine basis, unincorporated dye terminators were removed using Bio-Rad Biospin 30 columns, packed with 0.8 ml Bio-Gel® P-30 poly-acrylamide gel matrix in SSC, pH 8.0 (0.15 M sodium chloride, 17.5 mM sodium citrate) with 0.02 % sodium azide, which provide efficient non-interactive size separation. The column was inverted several times to resuspend the resin, the excess buffer was drained by gravity and discarded. The column and collection tube were centrifuged for 2 minutes in a swinging bucket rotor centrifuge at an rpm that generated a centrifugal force of 1,500 x g. The reaction was loaded to the centre of the resin, centrifuged for 4 minutes and the purified extension products collected in a fresh collection tube, precipitated with 150 µl 95 % ethanol and 5 µl 3 M sodium acetate, pH 5.2 and incubated on ice for at least 10 minutes (or may be stored at -20 °C overnight). The precipitated extension products were collected by centrifugation at 12,000 x g, washed with 200 µl 70 % ethanol (for an experienced user this step may be omitted but extreme care must be taken to ensure removal of all the ethanol/sodium acetate mix). The pelleted extension products were air dried (up to 10 minutes), resuspended in 6 µl 50 mM EDTA:deionised formamide (1 : 5) and 4 µl / 6 µl (36-well / 24-well comb respectively) loaded onto a pre electrophoresed (10 minutes) 6 % acrylamide / urea sequencing gel as described in the user manual.

Oligonucleotides designed for dye-deoxy™ terminator DNA sequencing.

When designing oligonucleotides for automated DNA sequence analysis, a number of criteria had to be taken into account (refer to optimisation of DNA sequencing for details). Oligonucleotides that were originally designed for the PCR reactions were purified further and used as sequencing primers. Specific oligonucleotides were designed to obtain sequence data from the 5' flanking region of MT-like genes from *A. thaliana* and were also designed to gene walk along the genomic clones obtained from *N. tobaccum.*

Arabidopsis thaliana.

Primer 792 5' GCC GCA CTT GCA GCC AGA TCC 3'

Arabidopsis thaliana.

Primer 894 5' GCA TAT ATG CAG AAC AGG CCA 3'

Primer 902 5' CAA CCA ATG GGA ATG CCC AC 3'

Nicotiana tabacum.

Primer 606 5' TCT GCT GGC ACT GTC AC 3'

Primer 613 5' GTG ACA GTG CCA GCA GA 3'

These two primers were designed to obtain further sequence data from 3 genomic clones from N. tabacum (clones 1, 2 and 13).

Primer 639 5' GAG TTG GAT TAC TAG GTA GG 3'

Primer 671 5' GAG AAA GTC AGG ACC GGA 3'

These two primers were designed to obtain further sequence data from clone 13.

Primer 640 5' GGG TGA TGT TTG AAA TCA ATG 3'

Primer 672 5' ACA GAA TGG ATG TCC ACA AC 3'

These two primers were designed to obtain further sequence data from clone 1.

Computerised analysis of DNA Sequences.

DNA sequences were analysed by the ABI SeqEd and DNA Strider programme. Sequence alignments were carried out using the Gap algorithms (GCG), data base searches using FASTA and multiple sequence alignment using PILEUP (Devereux *et al* 1984). Protein data base searching was carried out using OWL (Bleasby amd Wootton, 1990).

CHAPTER 3

RESULTS

Isolation of homologues of $PsMT_A$, a putative pea metallothionein gene.

Isolation of other members of the pea gene family using the polymerase chain reaction.

Using primers designed from DNA sequences obtained from *P. sativum* and *Mimulus guttatus* [primer sets 1 (190 and 191); 2 (238 and 239); 3 (284 and 283)], (refer to pages 76-77) and using the $PsMT_A$ genomic clone (pλS2) as a positive control. The reaction products were separated by agarose gel electrophoresis and 2 products were initially amplified from Feltham First (FF) pea genomic DNA (figure 3).

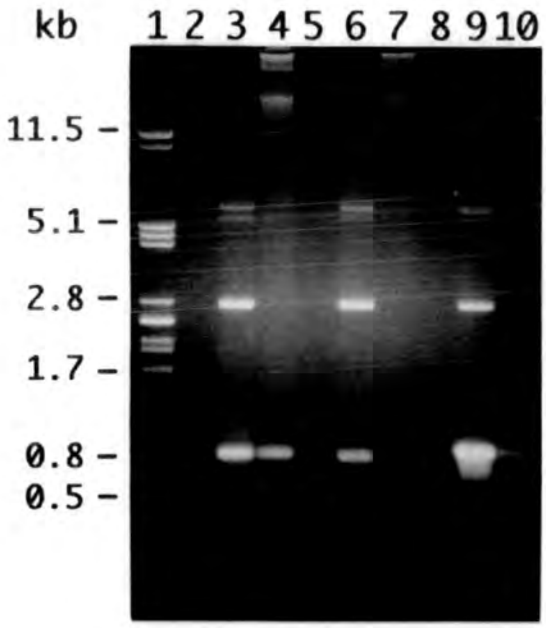
Primers	Template	Product size (bp)
190 and 191	pλS2 FF	805 805 and ca. 350
238 and 239	pλS2 FF	776 no products
284 and 283	pλS2 FF	825 825 and ca. 350

Table 1. PCR products amplified from pλS2 and FF genomic DNA.

Figure 3.

PCR using genomic clone pλS2 and pea genomic DNA with primer sets 1 (190 and 191), 2 (238 and 239) and 3 (284 and 283).

Lambda DNA digested with *Pst* I (lane 1), negative control using primer set 1 - no DNA (lane 2), positive control using primer set 1 - pλS2 (lane 3), pea genomic DNA using primer set 1 (lane 4), negative control using primer set 2 - no DNA (lane 5), positive control using primer set 2 - pλS2 (lane 6), pea genomic DNA using primer set 2 (lane 7), negative control using primer set 3 - no DNA (lane 8), positive control using primer set 3 - pλS2 (lane 9), pea genomic DNA using primer set 3 (lane 10).


kb
1 2 3 4 5 6 7 8 9 10
11.5
5.1
2.8
1.7
0.8
0.5

A second PCR reaction using primer sets 1 and 3 was prepared to obtain sufficient products for purification and subsequent cloning. Aliquots (20 μl) of the products were separated by agarose gel electrophoresis and a third product was observed of ca. 1 kb (figure 4). The products were transferred to nitrocellulose filter using Southern blotting and the filter probed with radiolabelled $PsMT_A$ cDNA. There was extensive hybridisation to all products and also hybridisation to a fourth product of ca. 1.2 kb amplified using primer set 3, which was not visualised by agarose gel electrophoresis and EtBr staining (figure 4).

The remainder of these products were resolved by agarose gel electrophoresis and the DNA electroeluted from gel slices, extracted with phenol/chloroform and precipitated with ethanol and sodium acetate (pH 5.2). T4 polymerase was used to create " blunt ends " on the PCR products and they were ligated to pUC 18 digested with *Hinc* II. DH5α sub-cloning efficiency competent cells were transformed to ampicillin resistance with the ligated DNA.

A number of white colonies were selected and DNA isolated using an alkaline lysis with PEG precipitation method (a modified version of the Birnboim and Doly alkaline lysis method, 1979). The inserts were released from the vector DNA by digestion with *Eco* RI and *Hind* III. The digested DNA was separated by agarose gel electrophoresis, transferred to nitrocellulose filter and probed with radiolabelled $PsMT_A$ cDNA. A number of the clones contained inserts of the expected size (ca. 800 bp, 350 bp and 1000 bp and designated $PsMT_A$, $PsMT_B$ and $PsMT_C$ respectively). Two clones for each product were sequenced using M13 forward and M13 reverse primers. The sequence data was compared to the sequence of the genomic clone pλS2 (figure 26).

Figure 4.

A. PCR using pea genomic DNA with primer set 1 (190 and 191) and primer set 3 (284 and 283).

B. Products were probed with radiolabelled *PsMT$_A$* cDNA which was generated via PCR using primer set 3 (284 and 283).

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), negative control using primer set 1 - no DNA (lane 2), pea genomic DNA using primer set 1 (lane 3), negative control using primer set 3 - no DNA (lane 4), pea genomic DNA using primer set 3 (lane 5).


A
kb
1 2 3 4 5
21.2 -
5.1 -
2.0 -
1.3 -
0.8 -
0.5 -
B
kb
1 2 3 4 5
21.2 -
5.1 -
2.0 -
1.3 -
0.8 -
0.5 -

Ligation-Mediated or Anchored Polymerase Chain Reaction (APCR).

As an alternative to screening genomic libraries for obtaining the 5' or 3' flanking regions of individual genes, an attempt was made to use ligation-mediated PCR to obtain the 5' flanking regions of $PsMT_B$ and $PsMT_C$. It was decided to use $PsMT_A$ in order to test the approach, since the 5' flanking region of this gene had already been characterised.

Vector DNA, pBluescript SK+ (5 µg), was digested with *Asp* 718 and purified using silica fines. The vector DNA was digested a second time with *Eco* RV and again purified using silica fines. Pea genomic DNA (10 µg) was digested with *Eco* RV and ligated to the purified vector in a 2:1 ratio (1.8 µg genomic DNA : 0.9 µg vector DNA). Anchored PCR (APCR) was performed on the ligated DNA using the generic M13 forward and M13 reverse primers and a specific primer designed to the intron of $PsMT_A$ (primer 301). These reaction products were separated by agarose gel electrophoresis but no products were observed.

A second attempt was made to use APCR to amplify the 5' flanking region of $PsMT_A$ but this time using size fractionated genomic DNA. Pea genomic DNA (100 µg) was digested with *Eco* RV (figure 5A), and size fractionated on a sucrose gradient (30-60 %). The sucrose gradient was centrifuged overnight and 1 ml fractions collected. Aliquots (200 µl) were loaded onto nylon membrane using a dot blot apparatus. The membrane was probed with a specific radiolabelled $PsMT_A$ PCR product (150 bp) which was amplified using primer 284 and the specific $PsMT_A$ primer 301. Eight fractions (#12 - #19) showed weak hybridisation to the probe (figure 5B). DNA was extracted from these fractions and the DNA ligated to pBluescript SK+ vector digested as previously described.

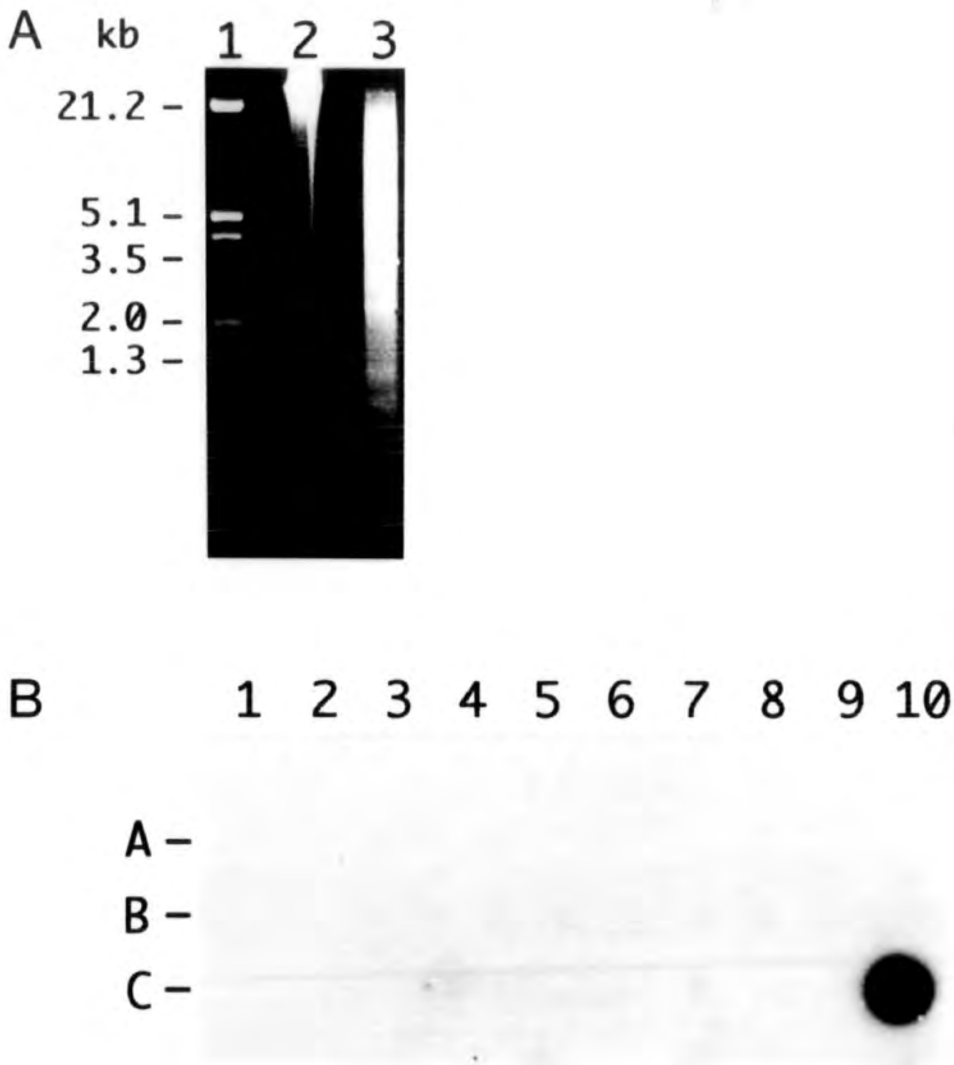
Figure 5.

A. Digestion of pea genomic DNA with *Eco* RV.

Panel A: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), pea genomic DNA undigested (lane 2), pea genomic DNA digested with *Eco* RV (lane 3).

B. Dot blot analysis of fractions from sucrose gradient and probed with probed with radiolabelled $PsMT_A$ PCR product generated using primers 284 and 301.

Panel B: Dot blot analysis of fractions collected from sucrose gradient; fractions 1-10 (A1-10), fractions 11-20 (B1-10), fractions 21-25 (C1-5), positive control $PsMT_A$ (C10).


A
kb
1 2 3
21.2 -
5.1 -
3.5 -
2.0 -
1.3 -
B
1 2 3 4 5 6 7 8 9 10
A -
B -
C -

APCR was performed using the ligated DNA as template (1 μl used in a 50 μl reaction) and an annealing temperature 50 °C. M13 forward and M13 reverse primers were used as the "generic primers" and primer 301 as the "specific primer". These reaction products were separated by agarose gel electrophoresis, the products transferred to nitrocellulose and the filters probed with radiolabelled $PsMT_A$ PCR product. Due to the specific ligation conditions, products should only be observed when using the M13 reverse primer and the 301 primer (488 bp). The expected products were not observed (figures 6 and 7).

The APCR reactions were repeated using more of the ligations as templates (5 μl used in a 50 μl reaction), the primer concentration remained the same but cycling parameters were altered by reducing the annealing temperature from 50 °C to 45 °C. These reaction products were separated by agarose gel electrophoresis and the products transferred to nitrocellulose and the filter probed with radiolabelled $PsMT_A$ PCR product. The results proved to be inconclusive as numerous fragments hybridised to the probe (figures 8 and 9).

This line of research was terminated (refer to discussion).

Figure 6.

A. APCR of DNA isolated from sucrose gradient fractions.

B. Products were probed with radiolabelled *PsMT$_A$* PCR product generated using primers 284 and 301.

Panels A and B: Lambda DNA digested with Pst I (lane 1), negative control - no DNA (lane 2), positive control - *PsMT$_A$* using primers M13 forward and reverse (lane 3), positive control - pea genomic DNA using primers 284 and 283 (lane 4), positive control - *PsMT$_A$* using primers 284 and 283 (lane 5), positive control - pea genomic DNA using primers 284 and 301 (lane 6), positive control - *PsMT$_A$* using primers 284 and 301 (lane 7), fraction 12 using primers M13 forward and 301 (lane 8), fraction 12 using primers M13 reverse and 301 (lane 9), fraction 13 using primers M13 forward and 301 (lane 10), fraction 13 using primers M13 reverse and 301 (lane 11).

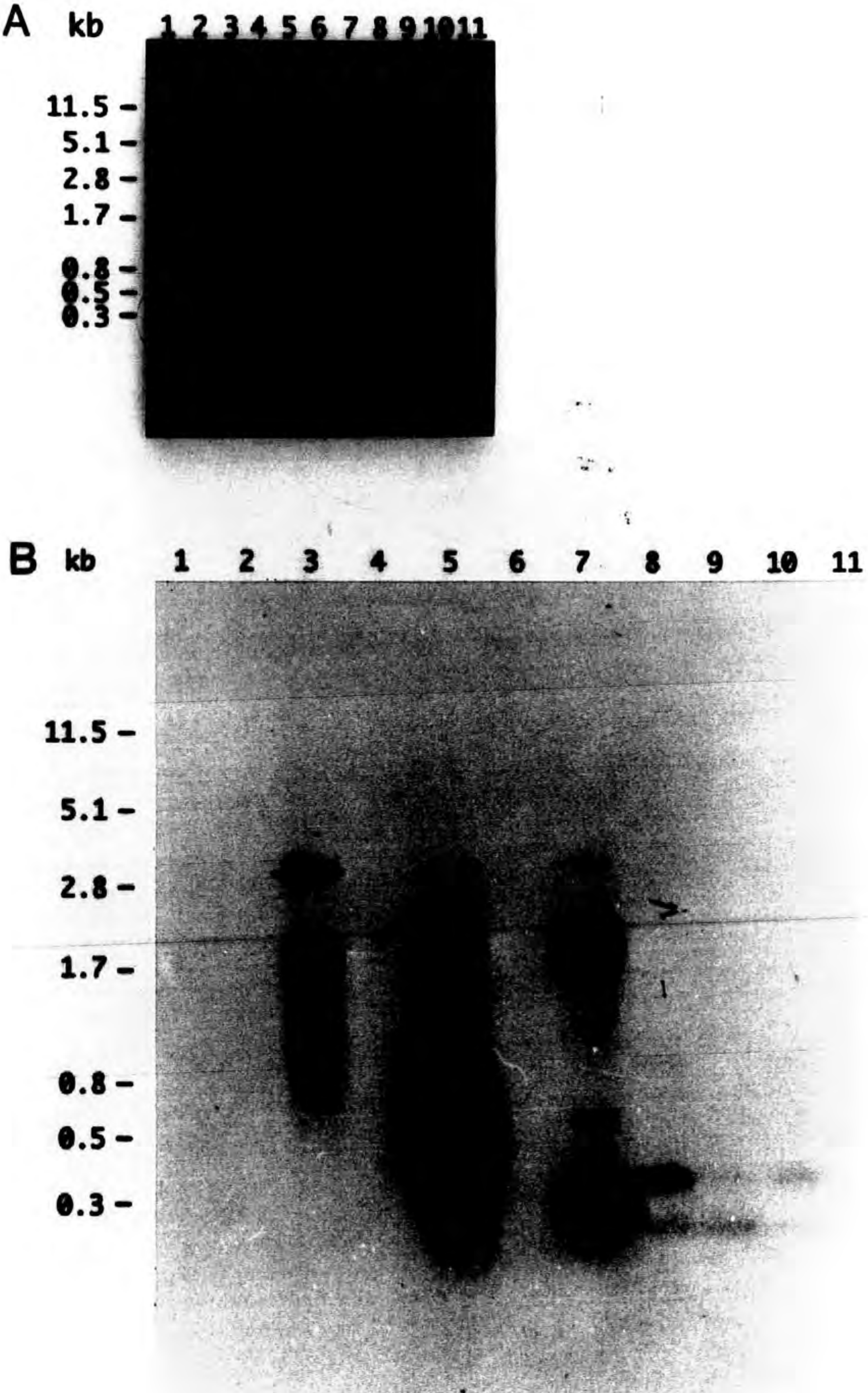

A
kb
1 2 3 4 5 6 7 8 9 10 11
11.5
5.1
2.8
1.7
0.8
0.5
0.3
B
kb
1 2 3 4 5 6 7 8 9 10 11
11.5
5.1
2.8
1.7
0.8
0.5
0.3

Figure 7.

A. APCR of DNA isolated from sucrose gradient fractions.

B. Products were probed with radiolabelled *PsMT$_A$* PCR product generated using primers 284 and 301.

Panels A and B: Lambda DNA digested with Pst I (lane 1), fraction 14 using primers M13 forward and 301 (lane 2), fraction 14 using primers M13 reverse and 301 (lane 3), fraction 15 using primers M13 forward and 301 (lane 4), fraction 15 using primers M13 reverse and 301 (lane 5), fraction 16 using primers M13 forward and 301 (lane 6), fraction 16 using primers M13 reverse and 301 (lane 7), fraction 17 using primers M13 forward and 301 (lane 8), fraction 17 using primers M13 reverse and 301 (lane 9), fraction 18 using primers M13 forward and 301 (lane 10), fraction 18 using primers M13 reverse and 301 (lane 11), fraction 19 using primers M13 forward and 301 (lane 12), fraction 19 using primers M13 reverse and 301 (lane 13).

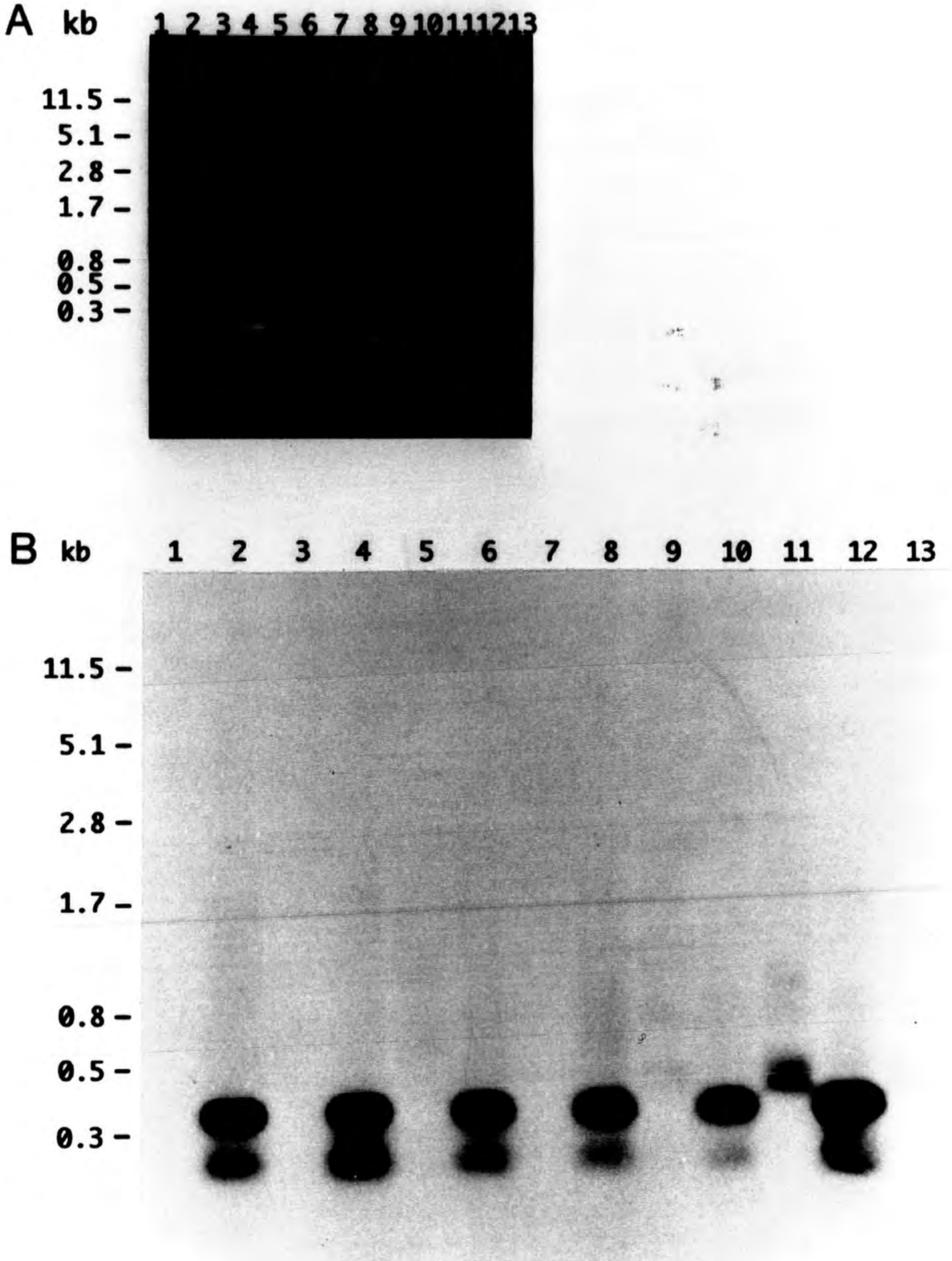

A
kb
1 2 3 4 5 6 7 8 9 10 11 12 13
11.5
5.1
2.8
1.7
0.8
0.5
0.3
B
kb
1 2 3 4 5 6 7 8 9 10 11 12 13
11.5
5.1
2.8
1.7
0.8
0.5
0.3

Figure 8.

A. APCR of DNA isolated from sucrose gradient fractions.

B. Products were probed with radiolabelled *$PsMT_A$* PCR product generated using primers 284 and 301.

Panels A and B: Lambda DNA digested with Pst I (lane 1), negative control - no DNA (lane 2), positive control - $PsMT_A$ using primers M13 forward and reverse (lane 3), positive control - pea genomic DNA using primers 284 and 283 (lane 4), positive control - $PsMT_A$ using primers 284 and 283 (lane 5), positive control - pea genomic DNA using primers 284 and 301 (lane 6), positive control - $PsMT_A$ using primers 284 and 301 (lane 7), fraction 12 using primers M13 forward and 301 (lane 8), fraction 12 using primers M13 reverse and 301 (lane 9), fraction 13 using primers M13 forward and 301 (lane 10), fraction 13 using primers M13 reverse and 301 (lane 11).

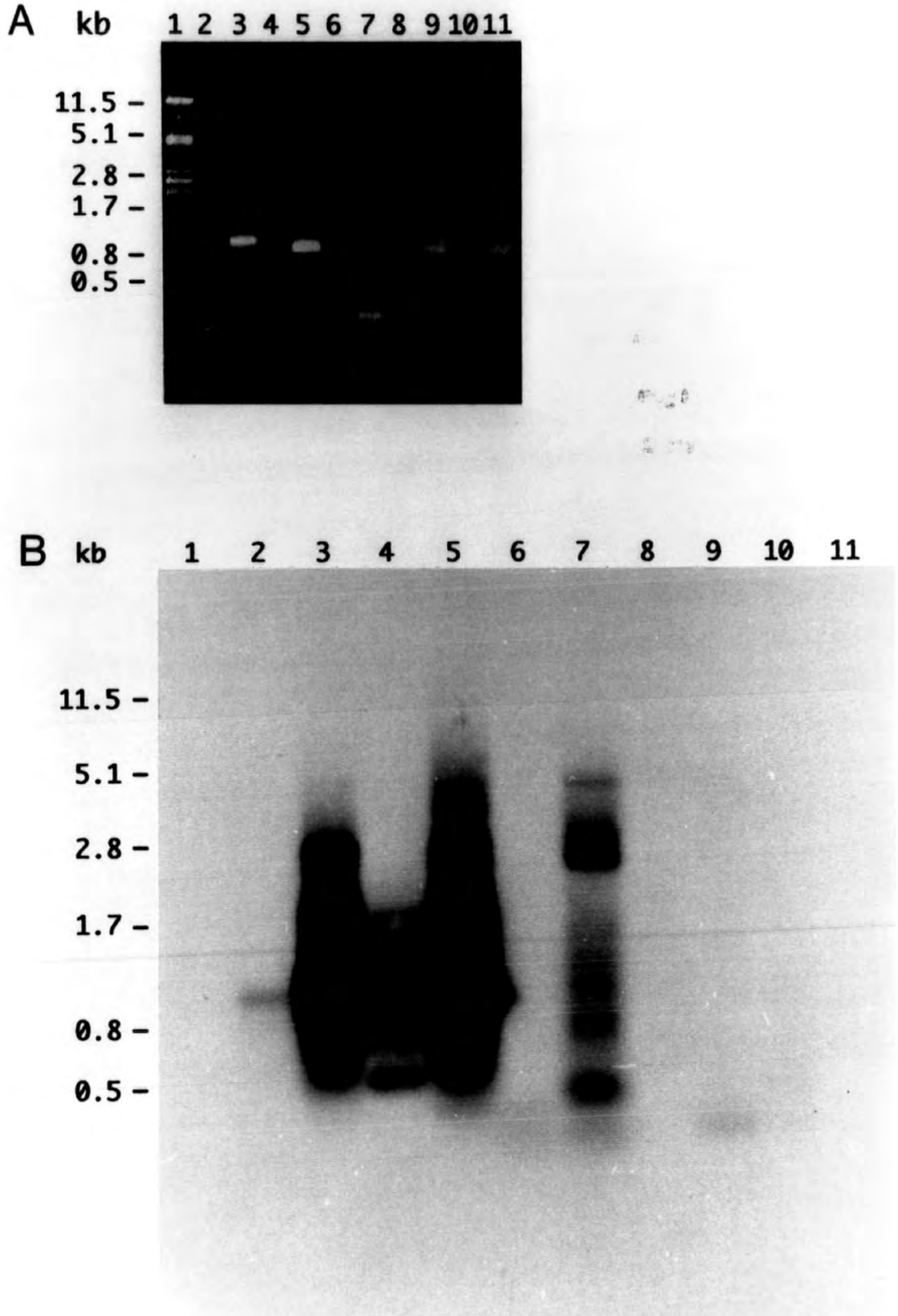
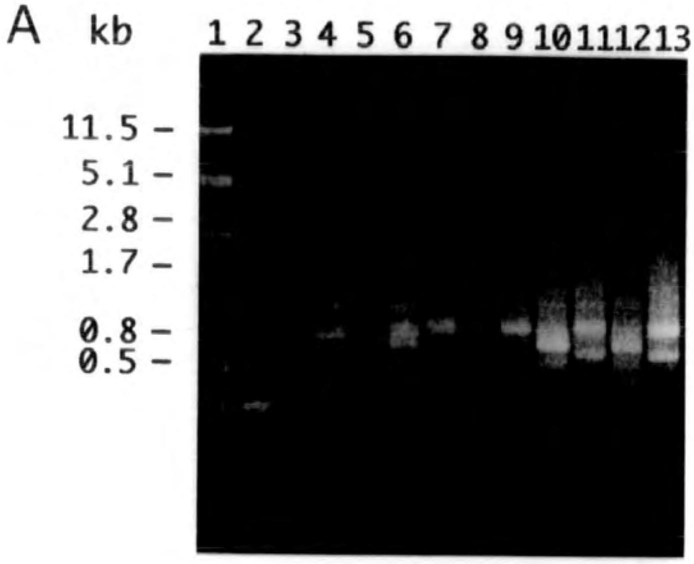

A
kb
1 2 3 4 5 6 7 8 9 10 11
11.5 -
5.1 -
2.8 -
1.7 -
0.8 -
0.5 -
B
kb
1 2 3 4 5 6 7 8 9 10 11
11.5 -
5.1 -
2.8 -
1.7 -
0.8 -
0.5 -

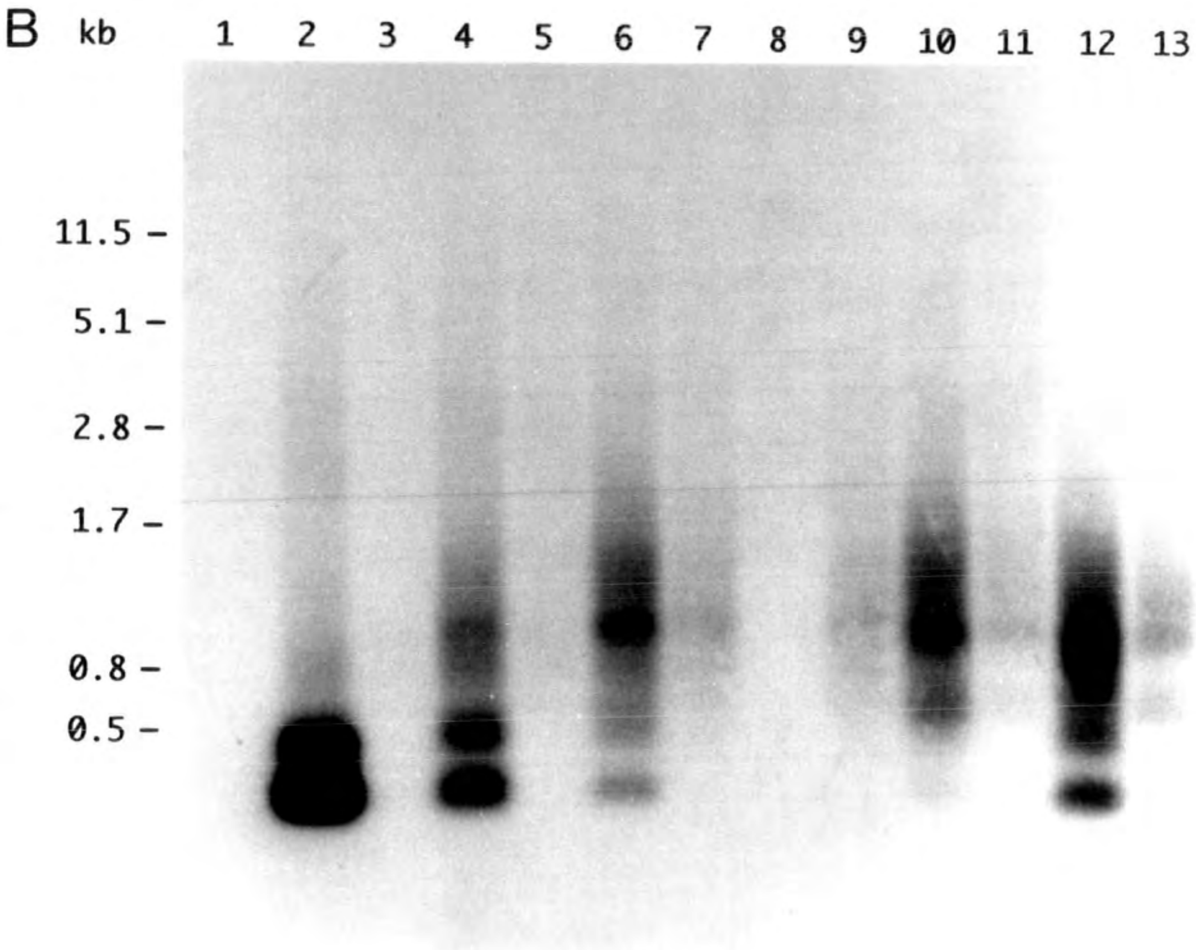
Figure 9.

A. APCR of DNA isolated from sucrose gradient fractions.

B. Products were probed with radiolabelled *PsMT$_A$* PCR product generated using primers 284 and 301.

Panels A and B: Lambda DNA digested with Pst I (lane 1), fraction 14 using primers M13 forward and 301 (lane 2), fraction 14 using primers M13 reverse and 301 (lane 3), fraction 15 using primers M13 forward and 301 (lane 4), fraction 15 using primers M13 reverse and 301 (lane 5), fraction 16 using primers M13 forward and 301 (lane 6), fraction 16 using primers M13 reverse and 301 (lane 7), fraction 17 using primers M13 forward and 301 (lane 8), fraction 17 using primers M13 reverse and 301 (lane 9), fraction 18 using primers M13 forward and 301 (lane 10), fraction 18 using primers M13 reverse and 301 (lane 11), fraction 19 using primers M13 forward and 301 (lane 12), fraction 19 using primers M13 reverse and 301 (lane 13).


A
kb
1 2 3 4 5 6 7 8 9 10 11 12 13
11.5 -
5.1 -
2.8 -
1.7 -
0.8 -
0.5 -


B
kb
1 2 3 4 5 6 7 8 9 10 11 12 13
11.5 -
5.1 -
2.8 -
1.7 -
0.8 -
0.5 -

Screening Pea Genomic Libraries.

Two pea genomic libraries were screened as previously described, in an attempt to isolate genomic clones corresponding to $PsMT_B$ and $PsMT_C$. The libraries were probed with radiolabelled $PsMT_A$ cDNA.

Library 1	(AMT)	primary screen	7 hybridising plaques identified
Library 2	(DPB)	primary screen	8 hybridising plaques identified

Plugs of each hybridising plaque were taken and subsequent secondary and tertiary screens were performed. Plugs from a number of single hybridising plaques were isolated from the tertiary screen and transferred to SM buffer and stored at 4 °C.

A PCR using lysate from the phage from library 1 with primer set 3 (284 and 283) was carried out. The reaction products were separated by agarose gel electrophoresis and transferred to nitrocellulose. A number of products were detected when probed with radiolabelled $PsMT_A$ cDNA (figure 10). Two products ca. 800 bp and 350 bp hybridised strongly to the probe from each reaction. The presence of more than one product in each reaction was not expected. The PCR reactions were repeated but this time the 350 bp product was not observed (data not shown).

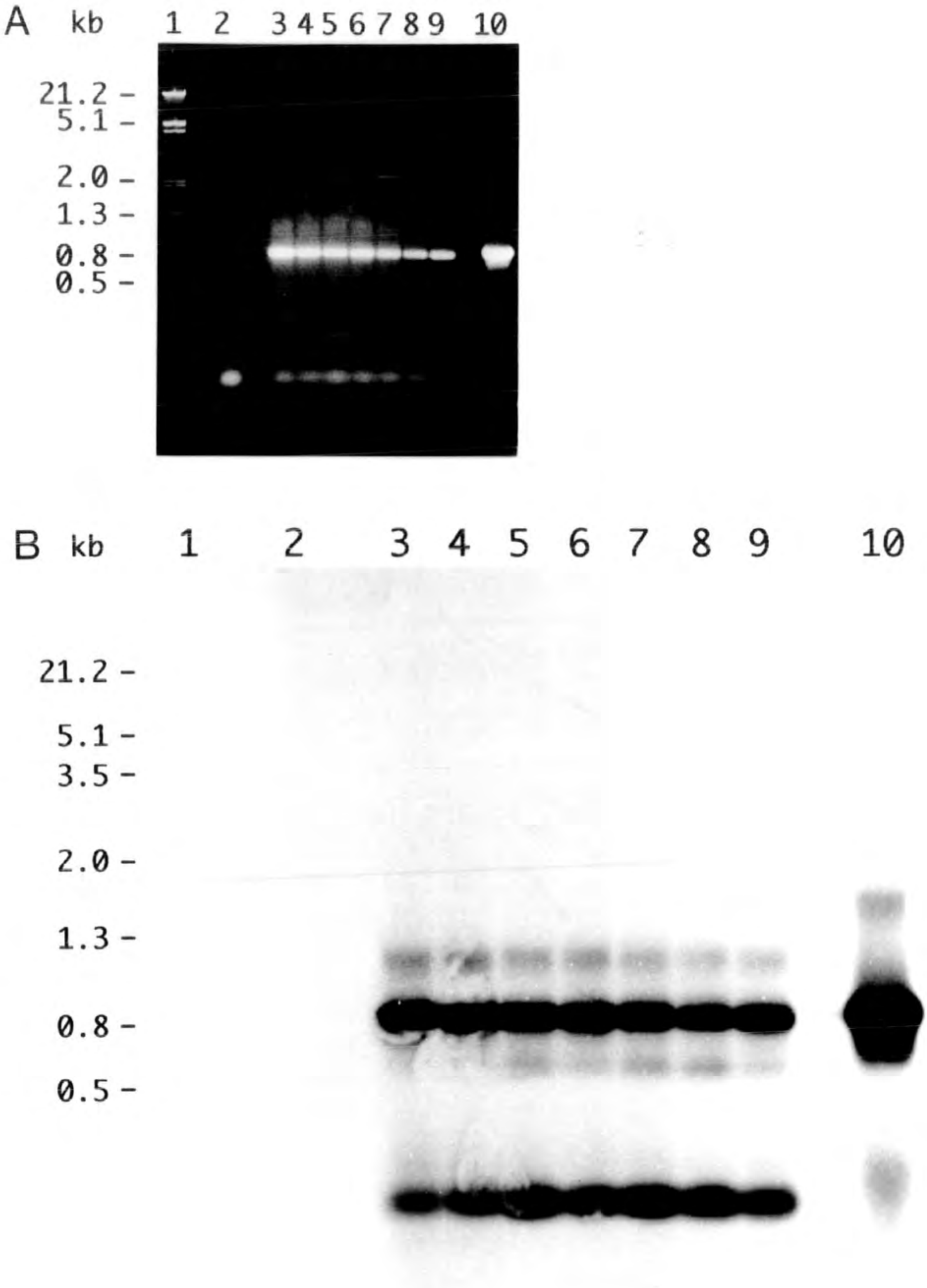
Due to problems encountered when trying to amplify specific sequences directly from phage, an attempt was made to isolate the phage DNA from plate lysates. Phage DNA isolated using a minipreparation procedure often resulted in very low yields with high levels of contaminating RNA. An attempt was then made to isolate lambda DNA using Qiagen Lambda DNA Purification Kit. Six hybridising plaques were selected, three from each library, but DNA was only isolated from hybridising plaques selected from library 2. The purified DNA was used in a PCR reaction with primer set 3 (284 and 283), and a product of ca. 820 bp was amplified from each DNA sample (data not shown).

Figure 10.

A. PCR of phage from library 1 using primers 284 and 283..

B. Products were probed with radiolabelled *$PsMT_A$* cDNA which was generated via PCR using primer set 3 (284 and 283).

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), negative control - no phage (lane 2), hybridising phage 1 (lane 3), hybridising phage 2 (lane 4), hybridising phage 3 (lane 5), hybridising phage 4 (lane 6), hybridising phage 5 (lane 7), hybridising phage 6 (lane 8), hybridising phage 7 (lane 9), positive control - *$PsMT_A$* (lane 10).

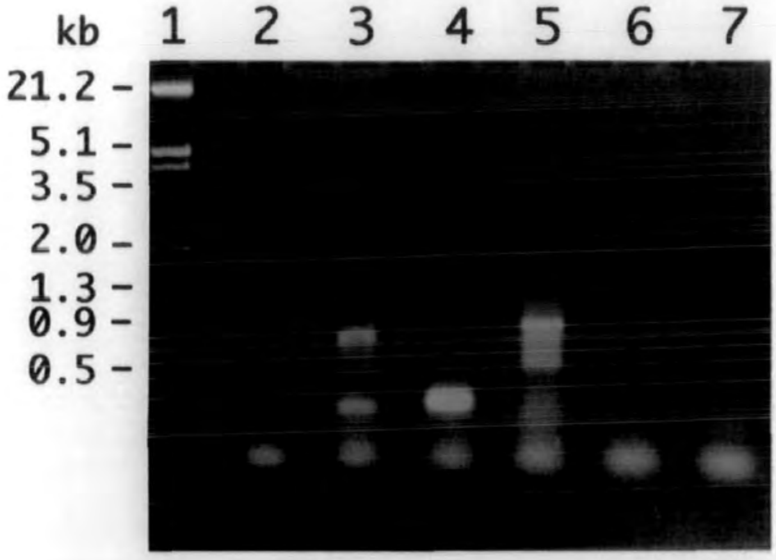

A
kb
1 2 3 4 5 6 7 8 9 10
21.2 -
5.1 -
2.0 -
1.3 -
0.8 -
0.5 -
B
kb
1 2 3 4 5 6 7 8 9 10
21.2 -
5.1 -
3.5 -
2.0 -
1.3 -
0.8 -
0.5 -

An attempt was made to amplify putative homologues from the original genomic libraries. Using primer set 3 (283 and 284), a PCR reaction was prepared with three positive controls: plasmid DNA from cloned $PsMT_A$, $PsMT_B$, $PsMT_C$ and experimental reactions from the two genomic libraries. Cycling conditions designed for amplifying products from libraries were used. The predicted products from the positive controls (825 bp, 350 bp and 1000 bp respectively) were observed when the PCR products were resolved by agarose gel electrophoresis. However, no products were observed for the two genomic libraries (figure 11).

Figure 11.

PCR of original pea genomic libraries using primer set 3 (284 and 283).

Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), negative control - no phage (lane 2), positive control - $PsMT_A$ (lane 3), positive control - $PsMT_B$ (lane 4), positive control - $PsMT_C$ (lane 5), pea genomic library 1 (lane 6), pea genomic library 2 (lane 7),


kb
1
2
3
4
5
6
7
21.2
5.1
3.5
2.0
1.3
0.9
0.5

Isolation of $PsMT_A$ homologues from other species.

Phaseolus vulgaris

A PCR reaction was prepared using 1.5 μg of *Phaseolus vulgaris* genomic DNA with primer sets 1-3. Pea genomic DNA was used as the positive control under the same reaction conditions. These reaction products were separated by agarose gel electrophoresis. A product of ca. 770 bp was observed with primer set 3. To increase the yield of the product from *P. vulgaris* a second PCR was prepared using different concentrations of primers (20 μM and 20 pM). The products were transferred to nitrocellulose filter and probed with radiolabelled $PsMT_A$ cDNA. There was strong hybridisation to the positive control and weak hybridisation to the product amplified from *P.vulgaris*. There was also hydridisation to the primer - dimer complex (figure 12).

The remainder of these products were resolved by agarose gel electrophoresis and the DNA electroeluted from gel slices, extracted with phenol/chloroform and precipitated with ethanol and sodium acetate (pH 5.2). The purified product was treated with T4 polymerase, to blunt the ends of the products, and ligated to pUC 18 restricted with *Hinc* II. DH5α sub-cloning efficiency competent cells were transformed to ampicillin resistance with the ligated DNA.

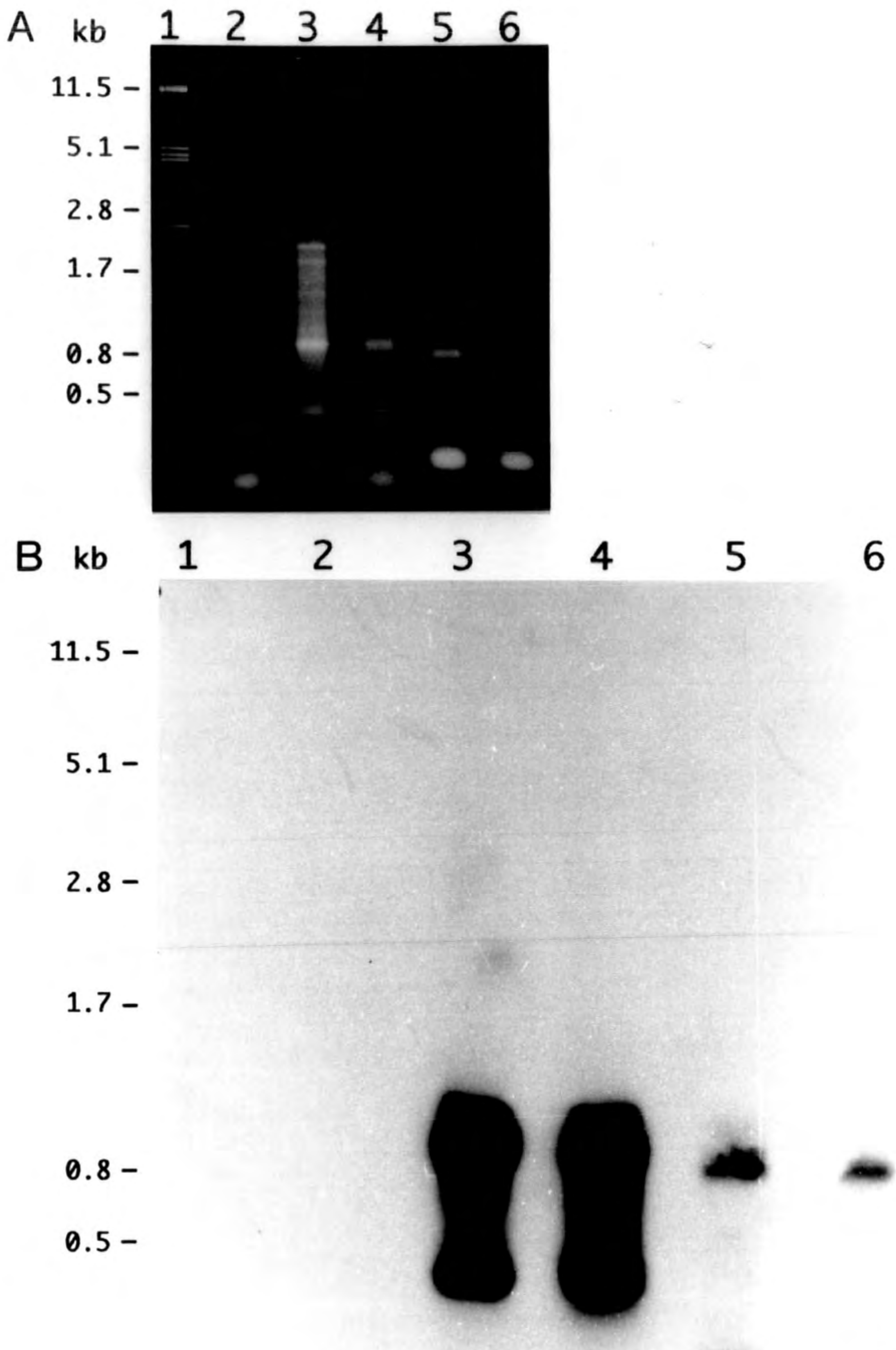
A number of white colonies were observed, therefore a colony screen was set up as previously described. The filter was probed with radiolabelled $PsMT_A$ cDNA. There was no hybridisation detected when the filter was exposed to x-ray film.

Figure 12.

A. PCR of *Phaseolus vulgaris* genomic DNA to amplify putative homologues of *PsMT$_A$* using primer set 3 (284 and 283).

B. Products were probed with radiolabelled *PsMT$_A$* cDNA which was generated via PCR using primer set 3 (284 and 283).

Panels A and B: Lambda DNA digested with *Pst* I (lane 1), negative control - no DNA (lane 2), positive control - pea genomic DNA (lanes 3 and 4), *Phaseolus vulgaris* genomic DNA (lanes 5 and 6).


A kb 1 2 3 4 5 6
11.5 –
5.1 –
2.8 –
1.7 –
0.8 –
0.5 –
B kb 1 2 3 4 5 6
11.5 –
5.1 –
2.8 –
1.7 –
0.8 –
0.5 –

Nicotiana tabacum

A partial cDNA sequence was obtained by screening a *Nicotiana tabacum* cDNA library with a $PsMT_A$ cDNA probe. This partial sequence was then used to screen a *N. tabaccum* genomic library. Three clones (1, 2 and 13) showed strong hybridisation to the *NtMT* cDNA probe. A series of nested deletions were performed on the 3 clones. DNA from the nested deletions were separated by agarose gel electrophoresis and transferred to nitrocellulose filter. The filters were probed with *NtMT* cDNA. From the autoradiograph, 18 of the nested deletions which hybridised to the probe were sequenced using the M13 forward primer.

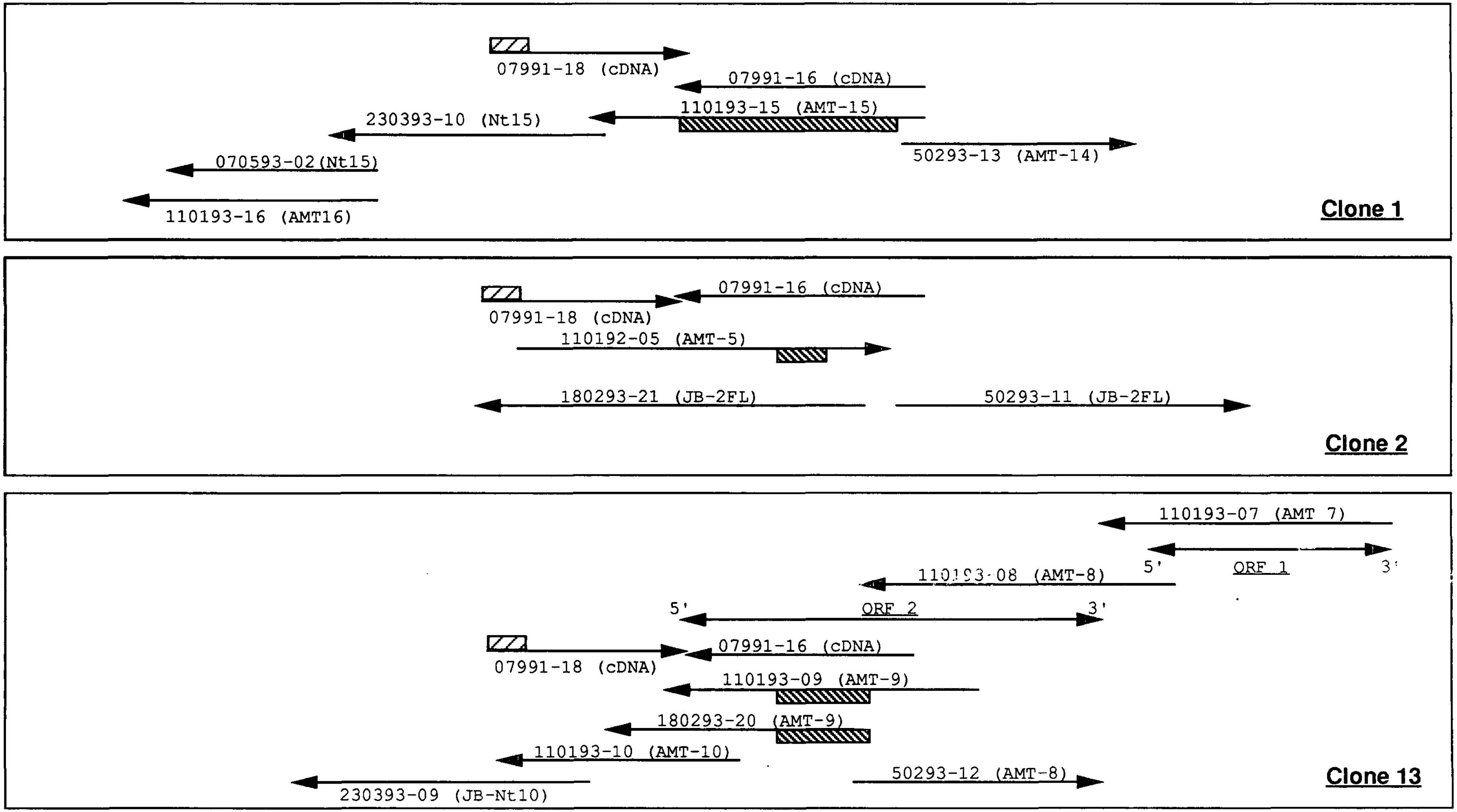
Clone 1	No. 15	280 base homology to *NtMT* cDNA.
Clone 2	No. 5	73 base homology to *NtMT* cDNA.
Clone 13	No. 13	120 base homology to *NtMT* cDNA.

Specific primers designed from the sequence data obtained with the M13 forward primer were synthesised to "gene walk" with the view to identify any Cys-Xaa-Cys motifs. Six primers were synthesised in total but subsequent sequencing failed to identify any regions encoding cysteine-rich motifs. However, clone 13 showed some sequence similarity to *Arabidopsis thaliana* meri 5 sequence (figure 28). Details of the sequencing performed is summarised in figure 13.

Figure 13. Sequencing strategy for genomic clones isolated from *N. tabacum*.

indicates the homologous regions of the *N. tabacum* genomic clones (1, 2 and 13) with respect to the *NtMT* cDNA.

indicates the region of the *NtMT* cDNA which encodes cysteine-rich motifs.


07991-18 (cDNA)
07991-16 (cDNA)
110193-15 (AMT-15)
230393-10 (Nt15)
50293-13 (AMT-14)
070593-02 (Nt15)
110193-16 (AMT16)
Clone 1
07991-16 (cDNA)
07991-18 (cDNA)
110192-05 (AMT-5)
180293-21 (JB-2FL)
50293-11 (JB-2FL)
Clone 2
110193-07 (AMT 7)
5'
ORF 1
3'
110193-08 (AMT-8)
5'
ORF 2
3'
07991-16 (cDNA)
07991-18 (cDNA)
110193-09 (AMT-9)
180293-20 (AMT-9)
110193-10 (AMT-10)
50293-12 (AMT-8)
230393-09 (JB-Nt10)
Clone 13

Arabidopsis thaliana MT-like gene.

Isolation and characteristion of *Arabidopsis thaliana MT-2* cDNA.

A PCR product of 247 bp was amplified from an *Arabidopsis thaliana* cDNA library using the primers 519 and 518 (restriction sites *Bam* HI and *Eco* RI were added to the 5' end of each primer respectively). The primers were designed to the original cDNA sequence (K. Takahashi, 1991) obtained from the EMBL database (accession number X628818). The product was isolated from an agarose gel using silica fines and cloned into pGEX 3X vector digested with *Bam* HI and *Eco* RI. The sequence was then determined using M13 forward and reverse sequencing. The sequence was confirmed to be *MT* -2 cDNA (figure 30).

The cDNA sequence in pGEX 3X was isolated by PCR using primers 519 and 518. The product was resolved using 1.0 % agarose gel electrophoresis and the DNA electroeluted from gel slices, extracted with phenol/chloroform and precipitated with ethanol and sodium acetate (pH 5.2). The concentration of the purified DNA was estimated via agarose gel electrophoresis with a known standard to be ca. 50 ng μl^{-1}. An aliquot, 6 µl (300 ng) of purified PCR product was radiolabelled using a Mega Prime Kit from Amersham as previouly described.

Isolation and characteristion of MT - like type 2 genomic sequence from *Arabidopsis thaliana*

An *Arabidopsis thaliana* genomic library was screened as previously described. Six hybridising colonies were identified from the primary screen when hybridised to an *MT* -2 cDNA probe. Secondary and tertiary screening were performed and two of the six hybridising colonies appeared to have been false positives.

Cosmid DNA was isolated from 9 hybridising colonies identified from the tertiary screen and designated clones 1-9 (which originated from the 4 hybridising colonies identified from the primary screen) using alkaline lysis miniprep method. The

DNA concentration was estimated via agarose gel electrophoresis to be ca. 0.5 $\mu g\ \mu l^{-1}$. Purified DNA (1.0 μg) was digested with *Hind* III and analysed on a 0.8 % agarose gel (figure 14). The digested DNA was transferred to nitrocellulose and the filter probed with radiolabelled *MT* -2 cDNA (figure 14).

An attempt was made to sequence the cosmid clones directly (refer to sequencing techniques for further details). Cosmid DNA was isolated using alkaline lysis with PEG precipitation and CsCl gradient as previously described. A minimum amount of sequence data was obtained from DNA prepared from an alkaline lysis miniprep method with the primer 519 used in the original PCR. Although the sequence data was ambiguous, there was sufficient data to confirm homology with the cDNA. A primer (792) was then designed to domain 1 of the cDNA to "gene walk" into the 5' flanking region of the gene. CsCl prepared cosmid DNA of genomic clones 2, 4 and 6 were used in sequencing reactions with primer 792 and dye-deoxy terminator chemistry and ca. 276 bases of 5' flanking region was obtained. The sequence data contained some ambiguities but the data confirmed that all three clones were identical and clone 4 has been designated *AtMT-2* (figure 30).

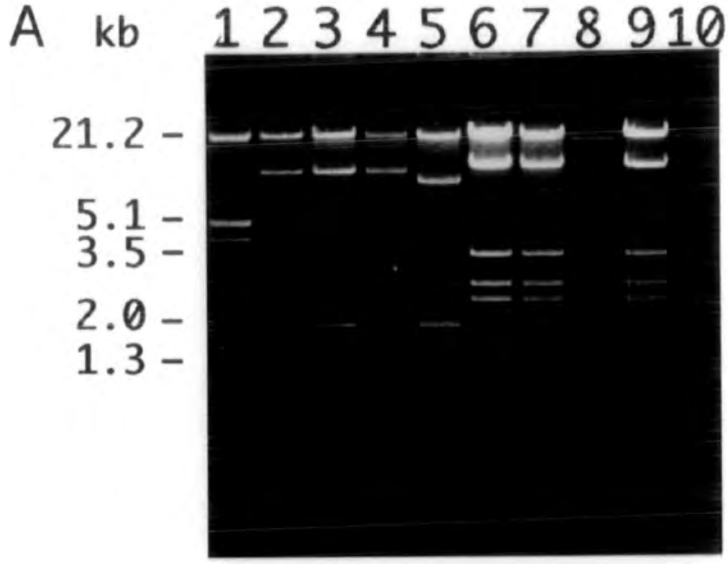
Due to the problems encountered when attempting to sequence cosmid DNA directly it was decided to try and subclone the *AtMT-2* gene. DNA (1.0 μg) from genomic clones 4 and 6 (concentration 0.5 and 0.75 $\mu g\ \mu l^{-1}$ respectively) were digested with 5 individual enzymes (*Bam* HI, *Eco* RI, *Hind* III, *Pst* I and *Sal* I). The digested DNA was separated by electrophoresis on a 0.8 % agarose gel and transferred to nitrocellulose and the filter probed with radiolabelled *MT -2* cDNA (figure 15). From the autoradiograph, the probe hybridised to one fragment in each track. The size of these fragments were in the range of ca. 7.4 kb - 14.5 kb. It was decided to prepare double digests of clone 4 to identify smaller fragments which would be more amenable to sub-cloning and subsequent DNA sequencing. DNA (1.0 μg) from genomic clone 4 was digested with (*Bam* HI and *Eco* RI, *Bam* HI and *Hind* III, *Bam* HI and *Pst* I,

Figure 14.

A. Digestion of *Arabidopsis thaliana* cosmid DNA with *Hind* III.

B. Digestions were probed with radiolabelled *AtMT* 2 cDNA which was generated via PCR using primer 519 and 518.

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), clones 1 - 9 digested with *Hind* III (lanes 2-10 respectively).


A
kb
1 2 3 4 5 6 7 8 9 10
21.2 -
5.1 -
3.5 -
2.0 -
1.3 -

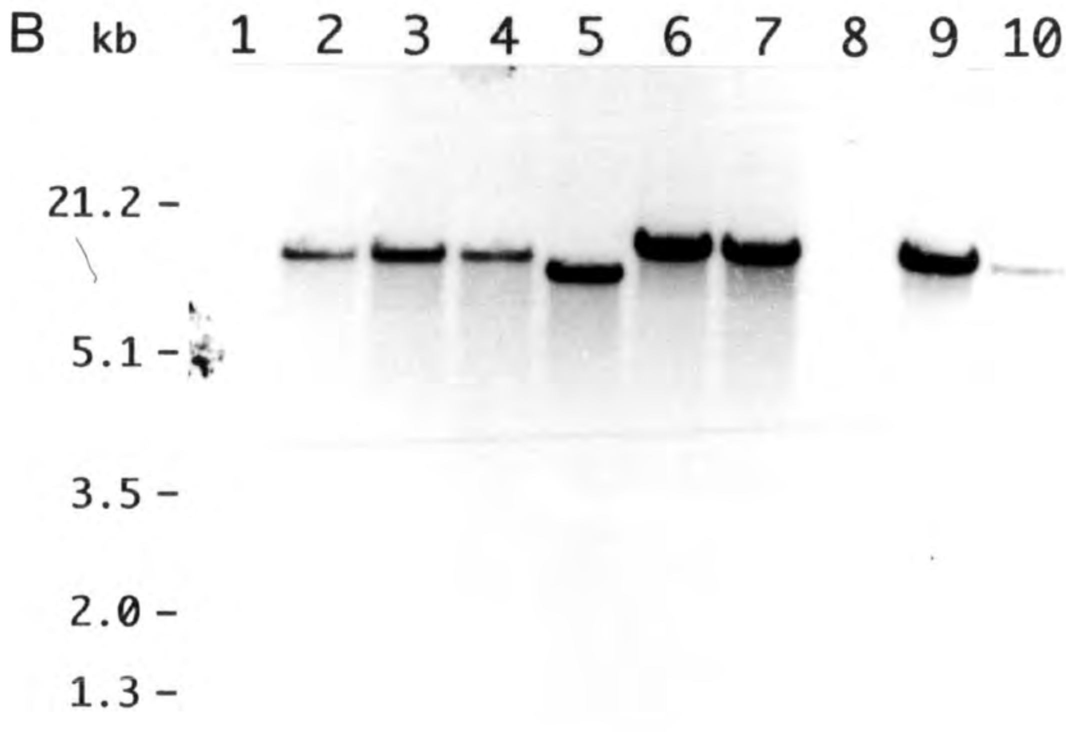
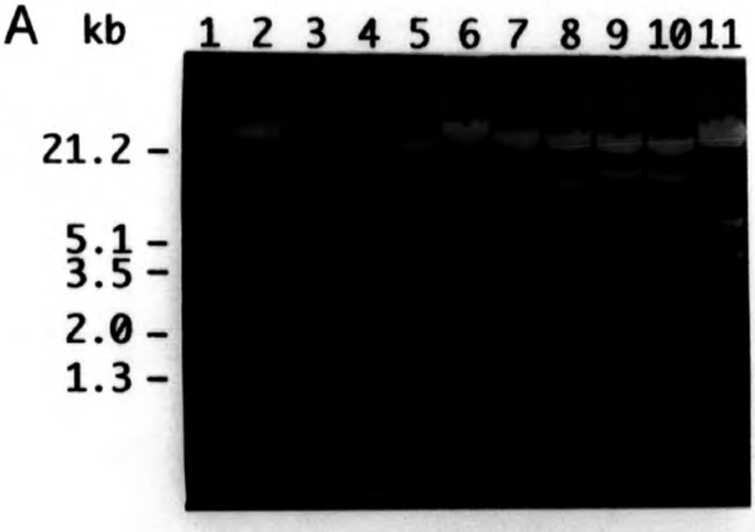

B
kb
1 2 3 4 5 6 7 8 9 10
21.2 -
5.1 -
3.5 -
2.0 -
1.3 -

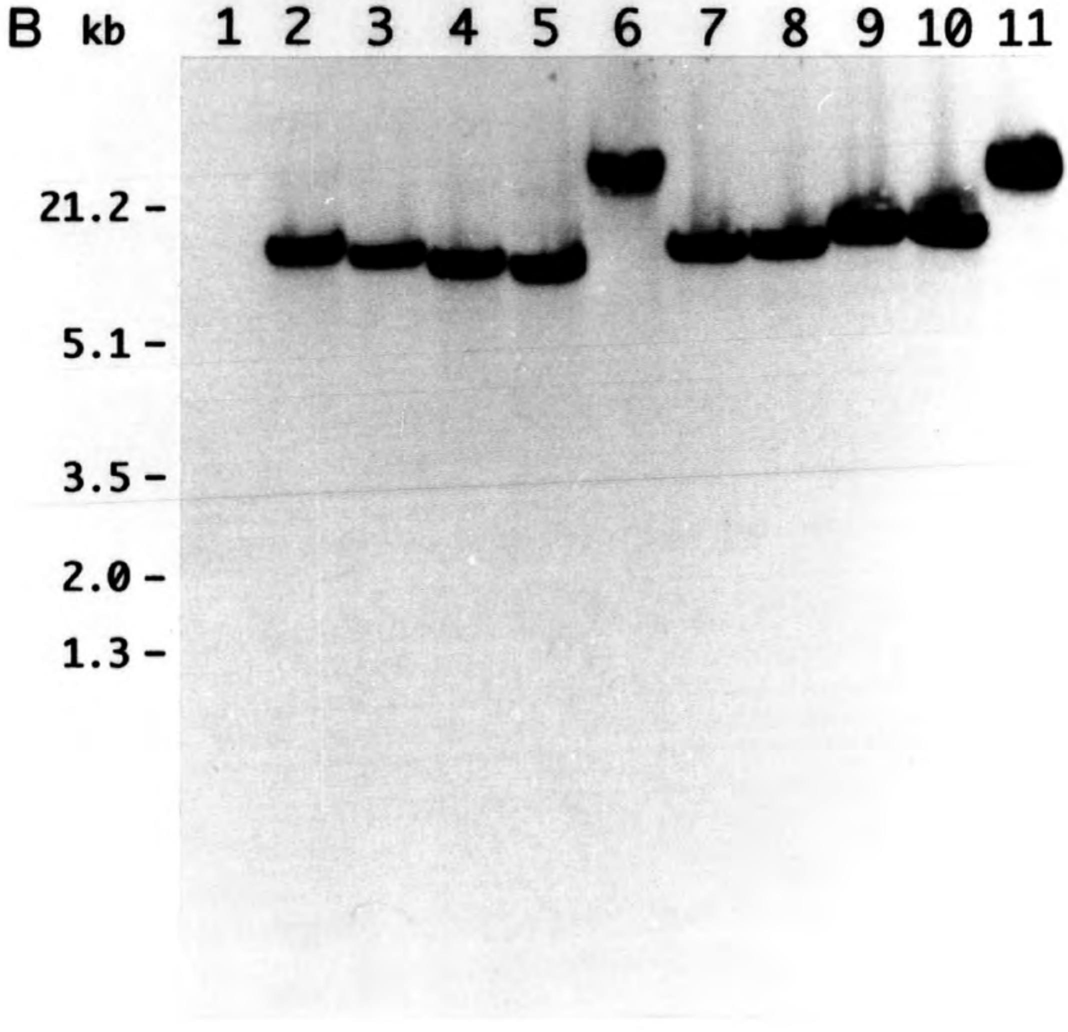
Figure 15.

A. Digestion of *Arabidopsis thaliana* cosmid DNA.

B. Digestions were probed with radiolabelled *MT* 2 cDNA which was generated via PCR using primer 519 and 518.

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), clone 4 digested with *Bam* HI, *Eco* RI, *Hind* III, *Pst* I, *Sal* I (lanes 2-6 respectively), clone 6 digested with *Bam* HI, *Eco* RI, *Hind* III, *Pst* I, *Sal* I (lanes 7-11 respectively)


A
kb
1 2 3 4 5 6 7 8 9 10 11
21.2
5.1
3.5
2.0
1.3


B
kb
1 2 3 4 5 6 7 8 9 10 11
21.2
5.1
3.5
2.0
1.3

Eco RI and *Hind* III, *Eco* RI and *Pst* I, *Hind* III and *Pst* I). The digested DNA was separated by electrophoresis on a 0.8 % agarose gel and transferred to nitrocellulose and the filter probed with radiolabelled *MT-2* cDNA (figure 16). Subsequent sub-cloning and screening of colonies failed to identify any clones with the correct size of insert.

Arabidopsis thaliana MT-like gene

Isolation and characteristion of *Arabidopsis thaliana* *MT-1* cDNA.

Using primers 826 and 827 which were designed to an *Arabidopsis thaliana* cDNA sequence, a PCR product of 295 bp was amplified from an *A. thaliana* cDNA library (figure 17A). The remainder of the PCR product was resolved using a 1.0 % agarose gel and the DNA isolated using a method from Zhen and Swank (1993). The DNA was extracted with phenol/chloroform, precipitated with ethanol and sodium acetate (pH 5.2) and the resulting DNA pellet resuspended in 10 μl sterile milli Q water. The DNA was quantified on gel and estimated to be ca. 50 ng μl^{-1}. To confirm the sequence of the amplified product, 2 μl (100 ng) was used in a sequencing reaction with primer 826 and dye-deoxy terminator chemistry. The sequence data confirmed that the amplified product was identical to *MT - 1* cDNA (Zhou and Goldsbrough, 1993). The signal strength of the sequence data diminished around base 200, this was probably due to too much DNA being added to the sequencing reaction (data not shown).

The PCR product (300 ng) was radiolabelled using a Mega Prime Kit from Amersham as previously described. A specific amount, 4.9 ng, of PCR product (295 bp) was ligated to 50 ng of pGem T vector (3 kb) achieving a 1:1 molar ratio. JM109 competent cells were transformed to carbenicillin resistance with the ligated DNA and the transformation plated onto LB agar containing x-gal and carbenicillin. A number of colonies were observed, therefore, a colony screen was set up and the filter probed with *MT -1* cDNA. The probe hybridised very strongly to the blue/white colonies (figure 17B). DNA was isolated from three of the hybridising colonies (5, 10 and 20) using a

Figure 16.

A. Digestion of *Arabidopsis thaliana* cosmid DNA.

B. Digestions were probed with radiolabelled *MT* 2 cDNA which was generated via PCR using primer 519 and 518.

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), clone 4 digested with *Bam* HI and *Eco* RI (lane 2), clone 4 digested with *Bam* HI and *Hind* III (lane 3), clone 4 digested with *Bam* HI and *Pst* I (lane 4), clone 4 digested with *Eco* RI and *Hind* III (lane 5), clone 4 digested with *Eco* RI and *Pst* I (lane 6), clone 4 digested with *Hind* III and *Pst* I (lane 7).

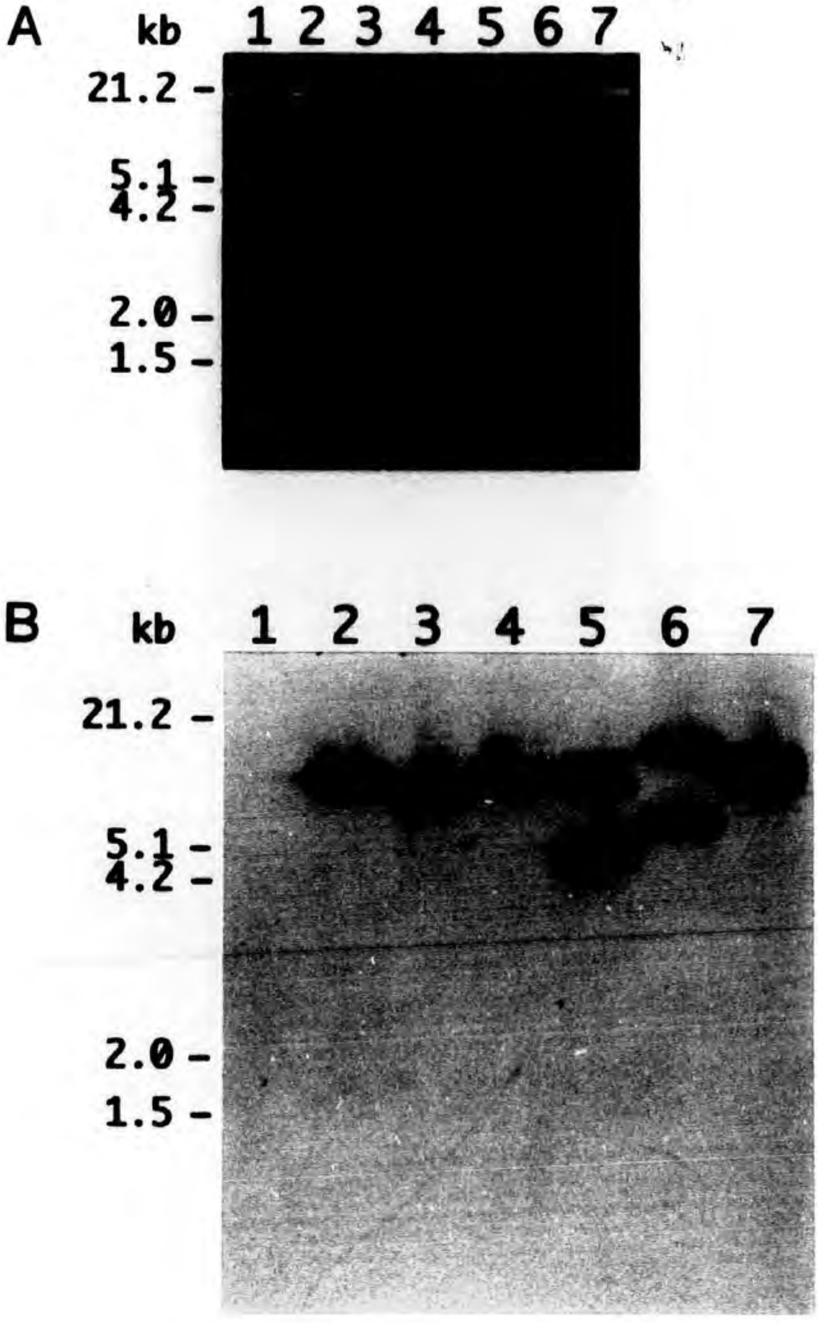

A
kb
1 2 3 4 5 6 7
21.2
5.1
4.2
2.0
1.5
B
kb
1 2 3 4 5 6 7
21.2
5.1
4.2
2.0
1.5

Figure 17.

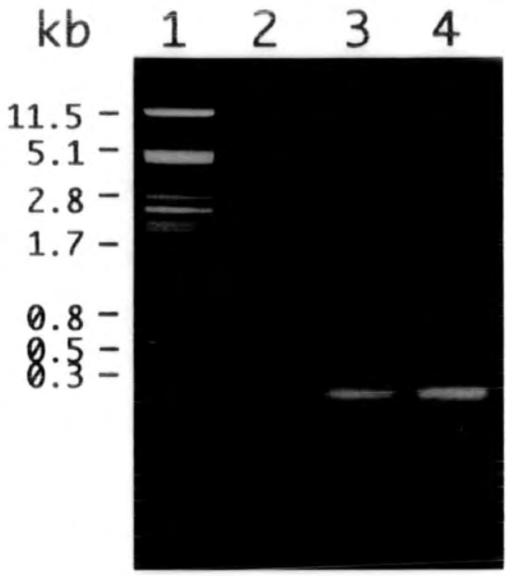
A. PCR of *Arabidopsis thaliana* cDNA using primers 826 and 827.

Panel A: Lambda DNA digested with *Pst* I (lane 1), negative control - no DNA (lane 2), *Arabidopsis thaliana* cDNA (lanes 3 and 4).

B. Cloning of purified PCR product into pGem T vector.

Panel B: Colony screen of PCR product cloned into pGem T vector probed with *MT-1* PCR product (295 bp).

A



B

modified version of the Promega Wizard minipreps and the inserts released by digesting the isolated DNA with *Nco* I and *Pst* I. The digested DNA was separated by agarose gel electrophoresis (figure 18A). The sequence was confirmed for one of the hybridising colonies (No. 20) using the M13 forward primer (figure 19).

Isolation and characteristion of MT - like genomic sequence from *Arabidopsis thaliana*

An *Arabidopsis thaliana* genomic library was screened as previously described. Six hybridising colonies were identified from the primary screen when hybridised to an *MT-1* cDNA probe. A secondary screening was performed and four of the six hybridising colonies appeared to have been false positives (figure 18B).

Cosmid DNA was isolated from the two hybridising colonies (designated A and B) identified from the secondary screen using an alkaline lysis miniprep method. The DNA concentration was estimated via agarose gel to be ca. 0.2 $\mu g\ \mu l^{-1}$.

Due to the problems already encountered when attempting to sequence cosmid DNA directly it was decided to subclone the *AtMT -1* gene. DNA (1.0 μg) from genomic clones A and B were restricted with 5 individual enzymes (*Bam* HI, *Eco* RI, *Hind* III, *Pst* I and *Sal* I). The digested DNA was run on a 0.8 % agarose gel and transferred to nitrocellulose and the filter probed with radiolabelled *MT-1* cDNA (figure 20). From the autoradiograph, the *MT-1* cDNA probe hybridised to one fragment in each track except for clone B digested with *Pst* I where there was hybridisation to three fragments. A suitably sized fragment was selected for sub-cloning. In this case a 3.5 kb *Pst* I fragment generated from clone A (lane 5) was selected.

Figure 18.

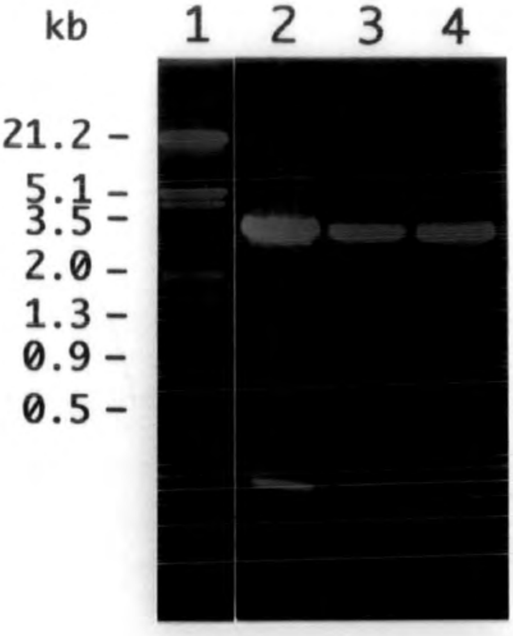
A. Digestion of DNA isolated from clones 5, 10 and 20 with *Nco* I and *Pst* I.

Panel A: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), clone 5 (lane 2), clone 10 (lane 3), clone 20 (lane 4).

B. Screening of *Arabidopsis thaliana* genomic library with *MT -1* cDNA generated via PCR using primers 826 and 827.

Panel B: Secondary screening of *Arabidopsis thaliana* primary screen, two hybridising colonies detected when probed with *MT-1* cDNA.

A



B

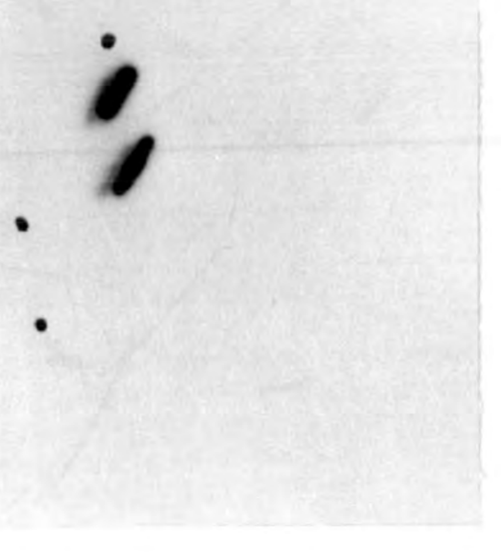


Figure 19.

Sequencing of cloned *MT -1* PCR product purified using Promega Wizard minipreparation (modified protocol).

Sequence data obtained using M13 forward primer and dye primer sequencing chemistry.

Applied Biosystems
Model 373A
Version 1.2.0

270194_32
DyePrimer{-21m13}
Lane 32
Signal: C:416 A:312 G:141 T:112

Points 1200 to [redacted] Jan 27, 1994 19:26
INSTRUMENT 901356
JB: 20C

X: 0 to 7115 Y: 0 to 1200
Spacing: 12.61

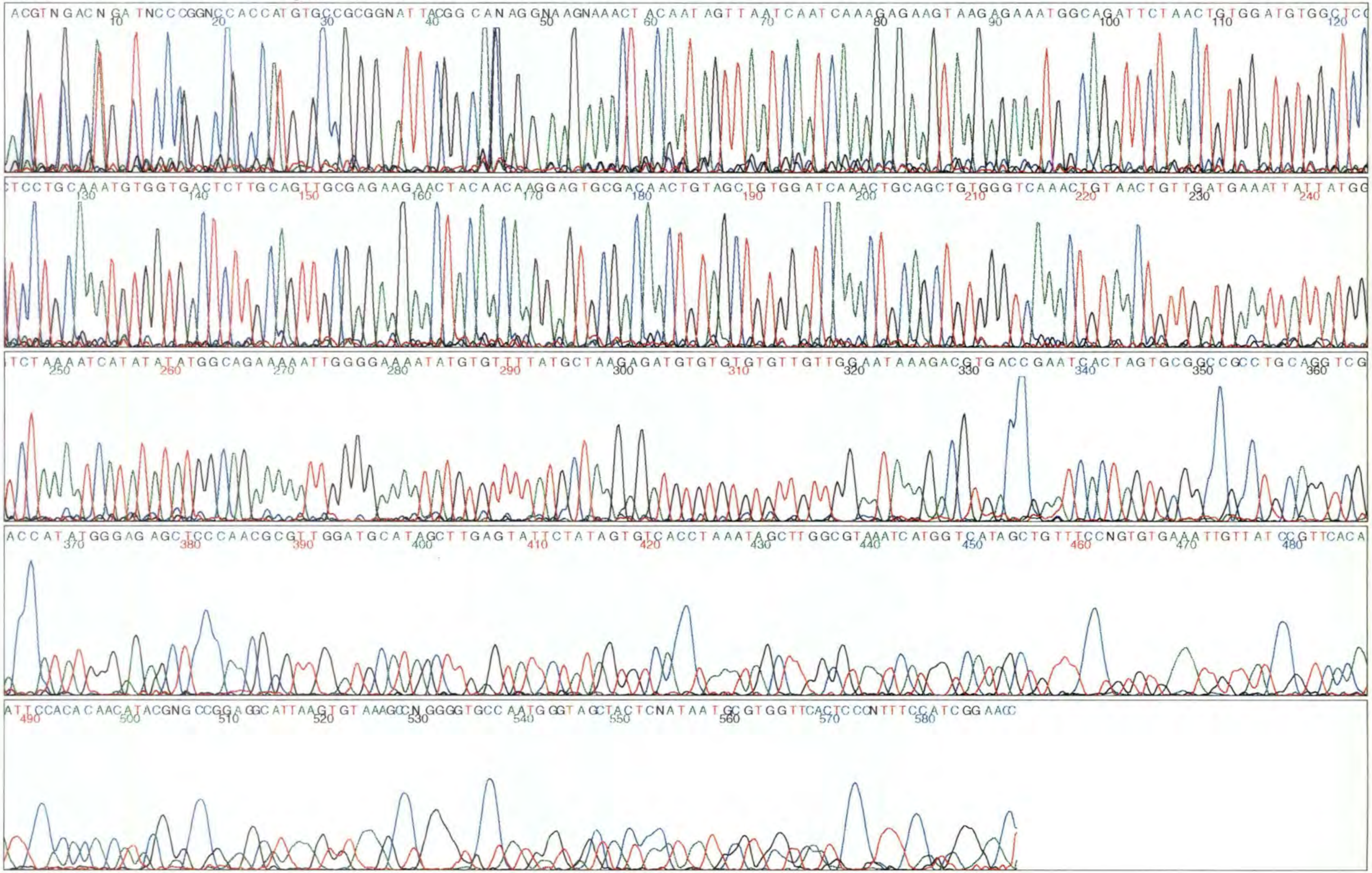
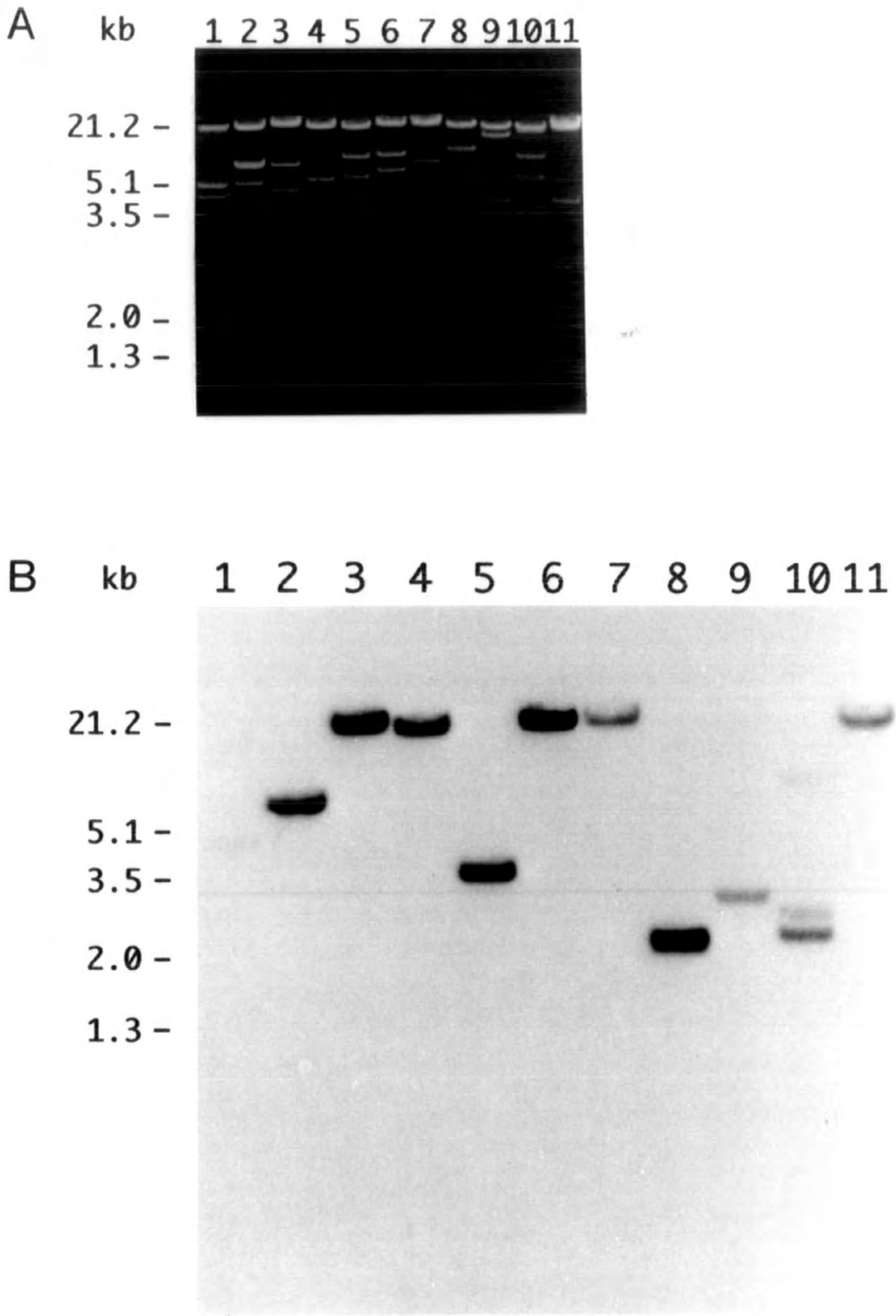


Figure 20.

A. Digestion of *Arabidopsis thaliana* genomic clones A and B.

B. Digestions were probed *MT-1* cDNA generated via PCR using primers 826 and 827.

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), clone A digested with *Bam* HI, *Eco* RI, *Hind* III, *Pst* I, *Sal* I (lanes 2 - 6 respectively), clone B digested with *Bam* HI, *Eco* RI, *Hind* III, *Pst* I, *Sal* I (lanes 7 - 11 respectively).


A
kb
1 2 3 4 5 6 7 8 9 10 11
21.2 -
5.1 -
3.5 -
2.0 -
1.3 -
B
kb
1 2 3 4 5 6 7 8 9 10 11
21.2 -
5.1 -
3.5 -
2.0 -
1.3 -

Cosmid DNA (2 x 2.5 μg) was digested with *Pst* I and separated using agarose gel electrophoresis. The 3.5 kb *Pst* I fragment was isolated from agarose gel using the method from Zhen and Swank (1993). The isolated DNA was ligated to pUC 18 digested with *Pst* I. DH5α sub-cloning efficiency competent cells were transformed to ampicillin resistance with the ligated DNA. Five white colonies were observed and DNA isolated from the white colonies and a blue colony for use as a control using a modified version of the Promega Wizard minipreps. The isolated DNA was digested with *Pst* I and separated using agarose gel electrophoresis. Undigested DNA (5.0 μl) was also separated by electrophoresis for quantification for sequencing purposes. One of the minipreps (no. 5) when digested with *Pst* I released an insert of the expected size (3.5 kb).

Prior to sequencing clone No.5, two test experiments were set up to confirm that the correct fragment had been cloned and to determine whether the fragment contained primer sites for primers 826 and 827 used to amplify the cDNA. The original cosmid DNA and the cloned 3.5 kb fragment were digested with *Pst* I and separated using a 0.8 % agarose gel. The DNA was transferred to nitrocellulose and the filter probed with radiolabelled *MT-1* cDNA. Hybridisation was observed for both cosmid and plasmid DNA confirming that the correct fragment had been isolated and cloned (figure 21).

A PCR reaction was prepared using the primers designed to the cDNA sequence of *MT -1*. One negative control, three positive controls and two experimental reactions were prepared. The products were separated using a 1.0 % agarose gel and transferred to nitrocellulose using a vacuum blotter and the filter probed with radiolabelled *MT -1* cDNA. There was extensive hybridisation to all of the products. Predicted products of 295 bp were detected for the cloned PCR product and the uncloned PCR product. By using PCR with the cosmid DNA and cloned 3.5 kb *Pst* I fragment, products of ca. 500 bp were observed. This gives us an indication that the gene for *AtMT -1* contains a small intron of ca. 200 bp (figure 22).

Figure 21.

A. Digestion of *Arabidopsis thaliana* genomic clone A and cloned 3.5 kb *Pst* I fragment with *Pst* I.

B. Digestions were probed *MT -1* cDNA generated via PCR using primers 826 and 827.

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), clone A digested with Pst I (lane 2), cloned 3.5 kb fragment digested with *Pst* I (lane 3).

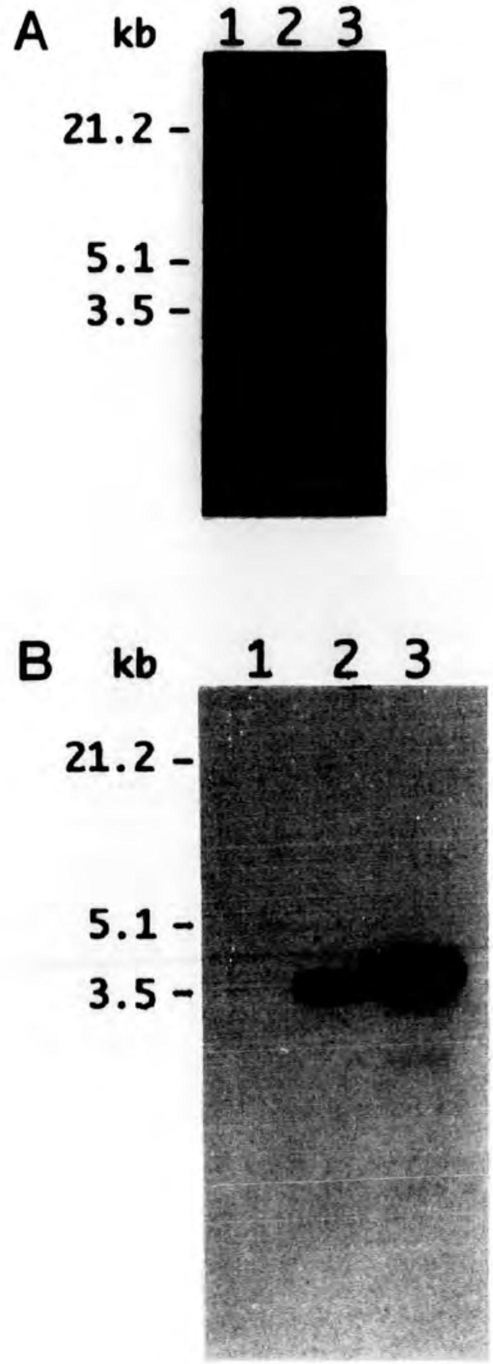
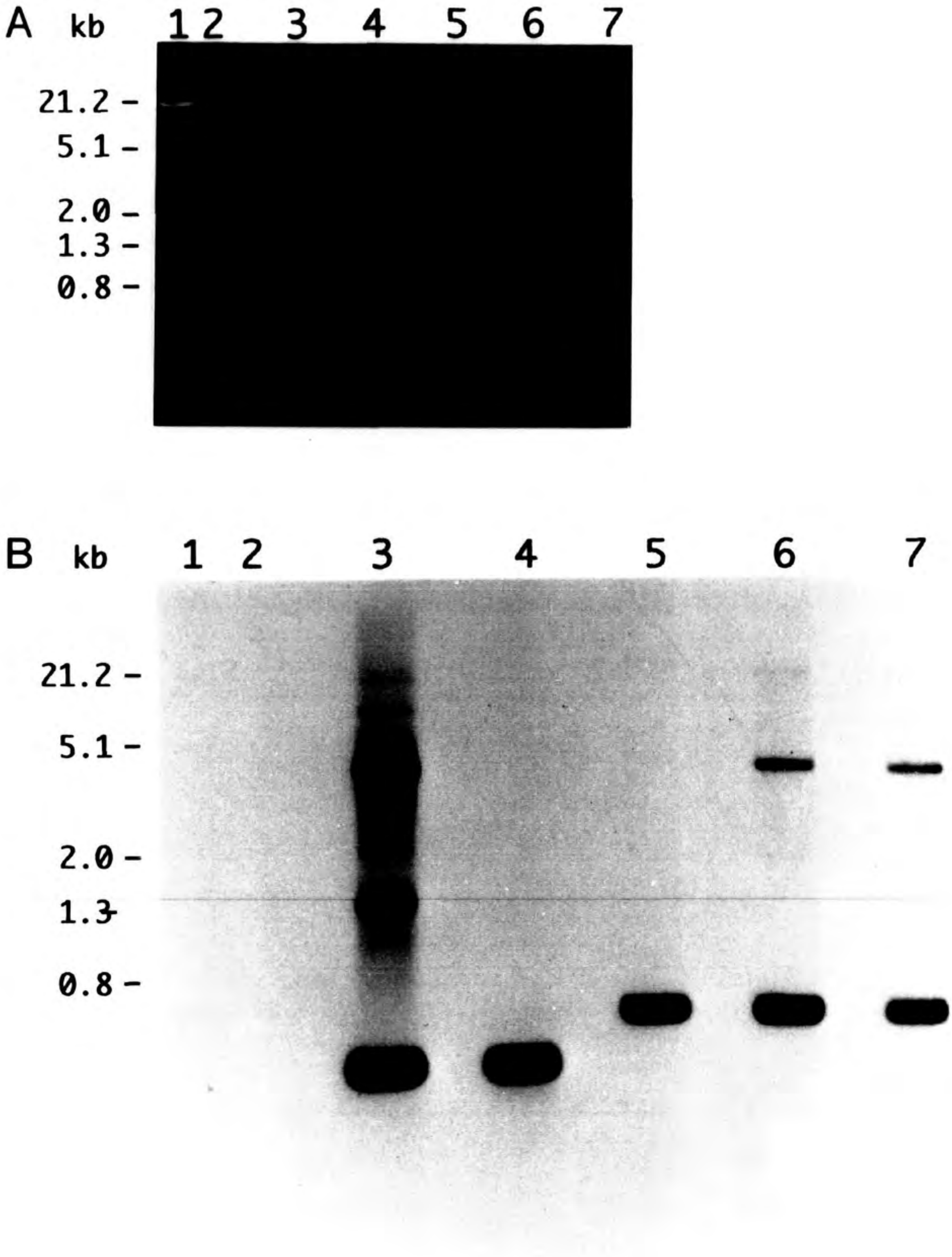

A
kb
1 2 3
21.2
5.1
3.5
B
kb
1 2 3
21.2
5.1
3.5

Figure 22.

A. PCR of *Arabidopsis thaliana* genomic clone A and cloned 3.5 kb *Pst* I fragment.

B. Products were probed *MT -1* cDNA generated via PCR using primers 826 and 827.

Panels A and B: Lambda DNA digested with *Eco* RI and *Hind* III (lane 1), negative control - no DNA (lane 2), positive control - cloned cDNA in pGem T vector (lane 3), positive control - purified cDNA from PCR (lane 4), positive control - cosmid clone A (lane 5), 3.5 kb *Pst* I fragment cloned into pUC 18 (lane 6), 3.5 kb *Pst* I fragment cloned into pUC 18 (lane 7).


A
kb
1 2 3 4 5 6 7
21.2 -
5.1 -
2.0 -
1.3 -
0.8 -
B
kb
1 2 3 4 5 6 7
21.2 -
5.1 -
2.0 -
1.3 -
0.8 -

The clone was then sequenced using M13 forward and M13 reverse primers. The sequence data obtained in both directions gave a high signal and low noise ratio. This again indicates the importance to obtain "pure" and "non-degraded" DNA at the recommended concentration for DNA sequencing (figure 23A and 23B). The DNA was subsequently sequenced with primers 826 and 827 using dye-deoxy terminator chemistry. The quality of the sequence data obtained with primer 827 was very good, however, no extension products were observed with primer 826. It was observed that a direct repeat is present in domain 2 of the cDNA nucleotide sequence (figure 29). When the sequence data of the genomic clone with primer 827 was analysed it was observed that one of these repeats had been completely deleted. A second sequencing reaction was performed and the same result was obtained.

Figure 23. DNA Sequencing of cloned 3.5 kb *Pst* I fragment.

A. Sequence data obtained using M13 universal primer and dye primer sequencing chemistry.

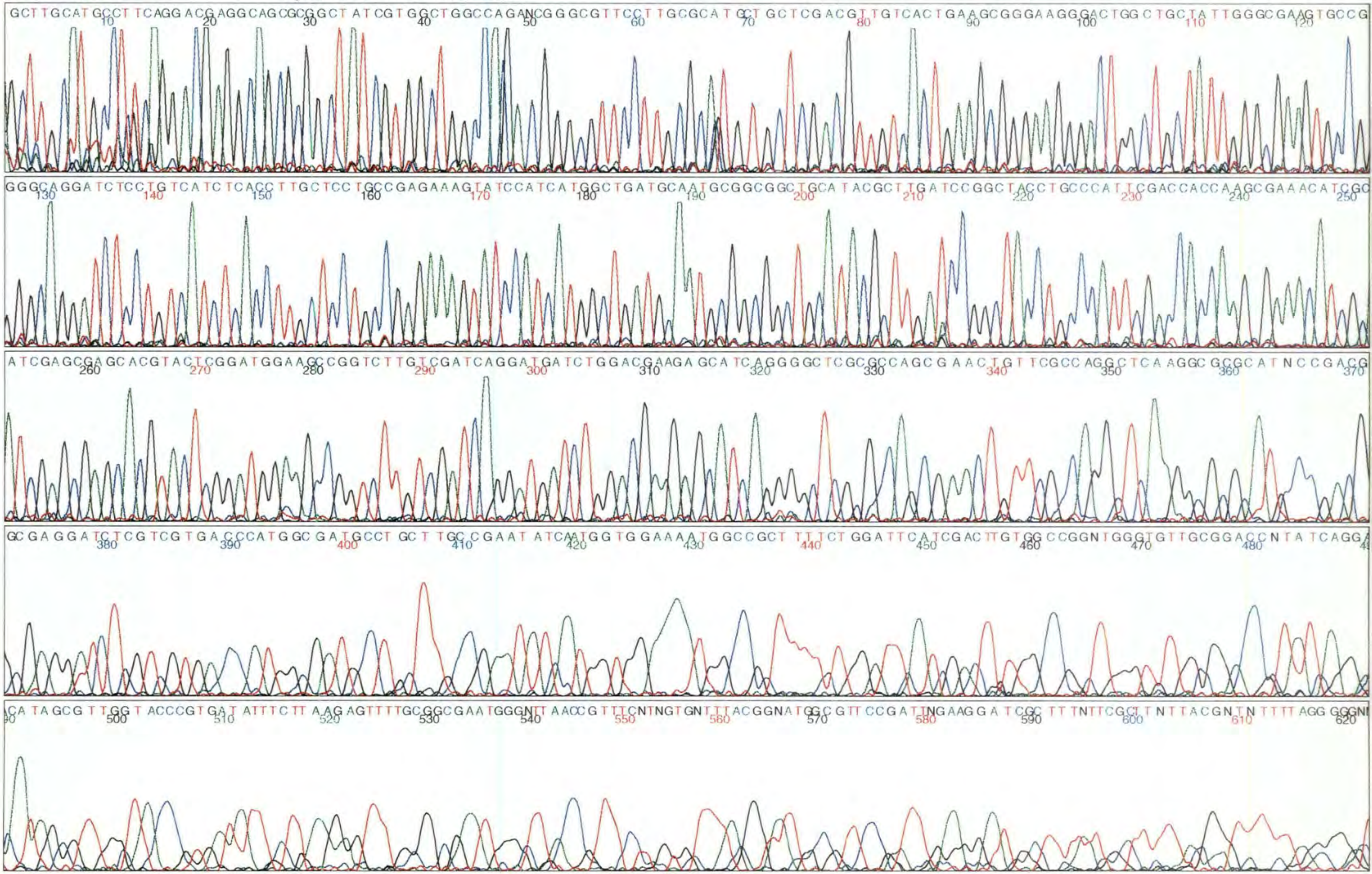
B. Sequence data obtained using M13 reverse primer and dye primer sequencing chemistry.

Applied Biosystems
Model 373A
Version 1.2.0

230294_16
DyePrimer{-21m13}
Lane 16
Signal: C:118 A:141 G:99 T:71

Points 810 to 7300 Base 1: 810
INSTRUMENT 901356
JB: pAtMT - 1 No. 5.

Wed, Feb 23, 1994 18:27
X: 0 to 7613 Y: 0 to 1200
Spacing: 11.37

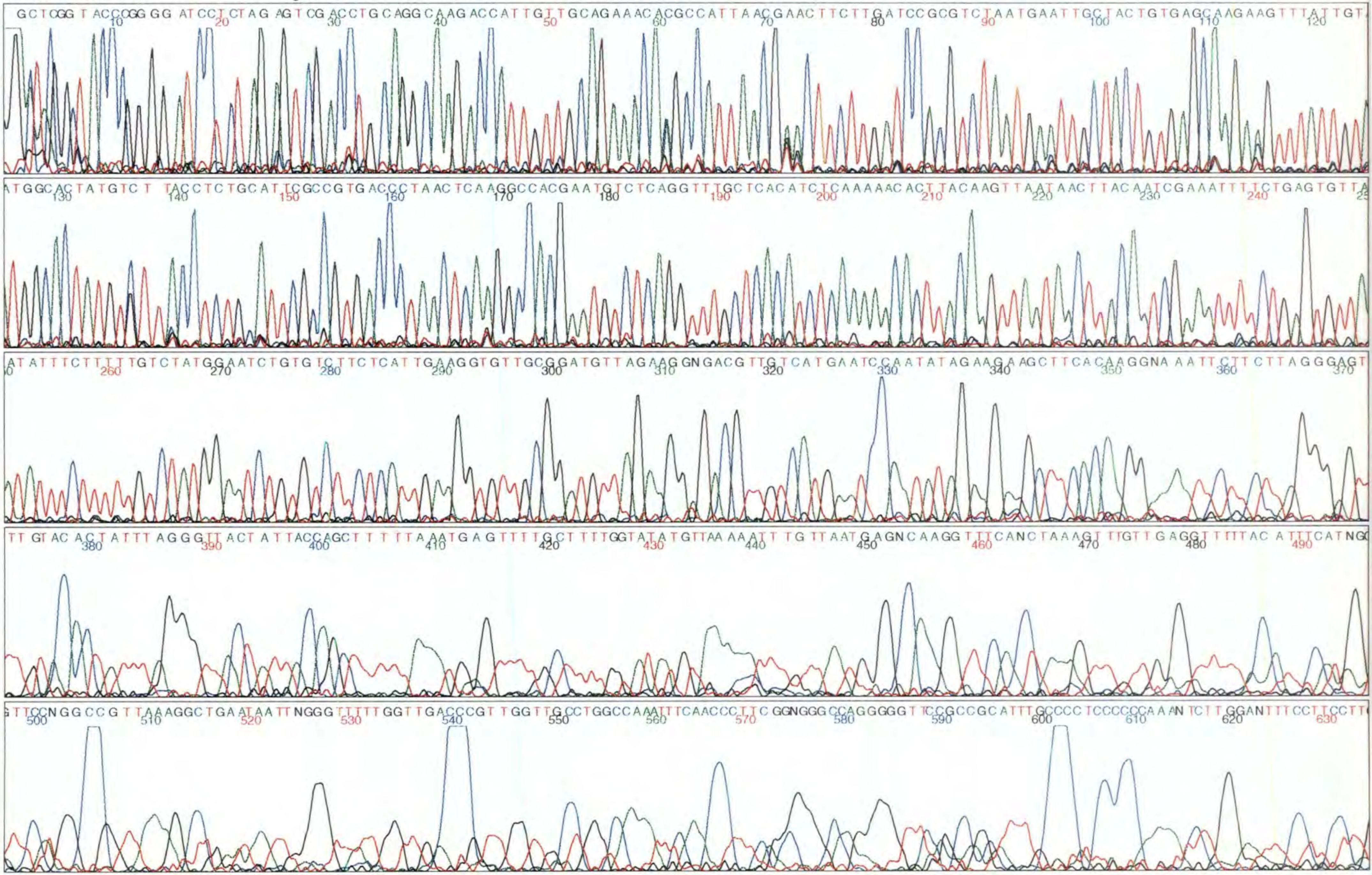


Applied Biosystems
Model 373A
Version 1.2.0

230294_08
DyePrimer{M13RP1}
Lane 8
Signal: C:172 A:135 G:60 T:50

Points 900 to 7500 Base 1: 900
INSTRUMENT 901356
JB: pAtMT - 1 No. 5.

Wed, Feb 23, 1994 18:27
X: 0 to 7681 Y: 0 to 1200
Spacing: 11.48



Arabidopsis thaliana MT-like gene

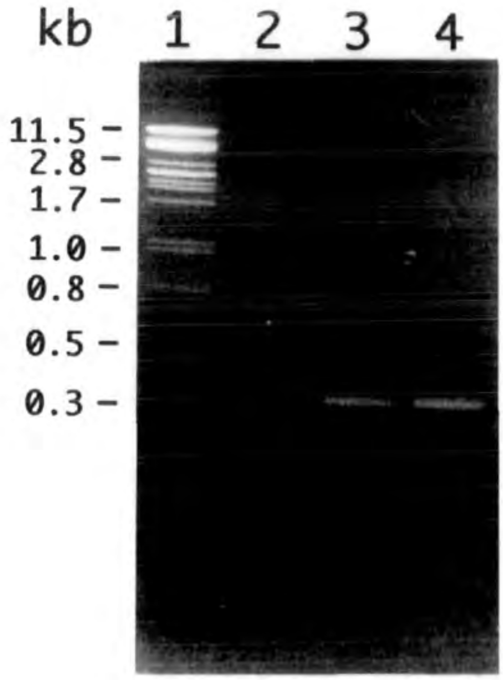
Isolation of *Arabidopsis thaliana* *MT- E_c* cDNA.

Using primers 900 and 901 which were designed to an *A.thaliana* sequence isolated from an *A.thaliana* seed cDNA library. Two 100 μl reactions were prepared, using 5 μl from an *A.thaliana* cDNA library as template, as previously described. Aliquots (20 μl) of the reaction products were separated by agarose gel electrophoresis. A weak fluorescent band of ca. 300 bp was visualised by uv from experimental reaction 2. However, there was insufficient product for isolation, therefore, two further PCR reactions were prepared using 20 μl from the first PCR reaction and aliquots (20 μl) of the reaction products were separated by agarose gel electrophoresis (figure 24). The remainder of the products were resolved using a 2.0 % agarose gel and the DNA isolated using a method by Zhen and Swank (1993). The DNA was extracted with phenol/chloroform, precipitated with ethanol and sodium acetate (pH 5.2) and the resulting DNA pellet resuspended in 10 μl sterile milli Q water. An aliquot (2 μl) of the DNA was quantified by agarose gel electrophoresis with a known standard and estimated to be < 20 ng μl^{-1}. Unfortunately, there was insufficient DNA to sequence the product directly.

Figure 24.

PCR of *Arabidopsis thaliana* cDNA using primers 900 and 901.

Lambda DNA digested with Pst I (lane 1), negative control - no DNA (lane 2), *Arabidopsis thaliana* cDNA (lanes 3 and 4).


kb
1
2
3
4
11.5
2.8
1.7
1.0
0.8
0.5
0.3

Optimisation of automated DNA sequence analysis.

During the course of this study, standard procedures have been optimised for use with the Applied Biosystems 373A DNA Sequencer and these are now routinely used for the successful running of the University of Durham DNA Sequencing Service.

Template preparation.

A number of template preparations have been under investigation to identify suitable methods to produce templates of a suitable standard for automated DNA sequence analysis. Methods for purifying plasmid, cosmid and PCR generated DNA have been optimised and the results detailed below.

Modified Birnboim and Doly (1979) alkaline lysis with PEG precipitation.

This method when followed correctly yields pure DNA for sequencing (yields may vary from as little as 3 μg up to 20 μg, however, yields of more than 20 μg were not uncommon). When the DNA concentration had been quantified correctly an individual sequencing reaction produced sequence data of 450-550 bases in length. It should also be noted that all three ethanol washes are required as these are very important for removing PEG which inhibits Taq polymerase in the sequencing reaction and the DNA resuspended in sterile milli Q water, not Tris buffer / TE to reduce salt concentration.

Promega Magic / Wizard Minipreps (improved protocol).

This method is a modification of the protocol described by the manufacturer. When using the protocol detailed by the manufacturer, the resulting DNA proved to be an unreliable subsrtate for automated sequence analysis. This is possibly due to either excess salt or guanadine hydrochloride which buffers the resin being present in the purified DNA sample. For some samples it was possible to improve the quality of the sequence data by precipitating the DNA with either ethanol or isopropanol prior to DNA sequencing. However, the adapted protocol yielded pure DNA for sequencing (yields may vary from as little as 5 μg up to 30 μg, most commonly the DNA concentration was 200-250 ng μl^{-1} which was ideal for double-stranded DNA sequencing). When the DNA concentration had been quantified correctly an individual sequencing reaction produced sequence data of 400-500 bases in length.

Qiagen QIAprep-spin Plasmid Kit.

This method when followed correctly yielded very pure DNA for sequencing (yields may vary from as little as 5 μg up to 30 μg, most commonly the DNA concentration was 200-250 ng μl^{-1} which was ideal for double-stranded DNA sequencing). When the DNA concentration had been quantified correctly an individual sequencing reaction produced sequence data of 450-550 bases in length.

Cosmid DNA prepared using the modified Birnboim and Doly alkaline lysis with PEG precipitation.

A number of attempts were made to sequence cosmid DNA purified using the modified Birnboim and Doly alkaline lysis with PEG precipitation method (small scale). An example of sequence data produced by this method was presented in an Applied Biosystems Research Reporter (Issue No. 14, November 1991). These attempts failed to produce cosmid DNA of a suitable quality or concentration for DNA sequencing. Recommended modifications (pers. comm. Applied Biosystems, Foster City, California, USA) for use with this protocol were carried out: Using a whirlimixer was avoided to prevent any shearing of the DNA, the resulting DNA was resuspended in 50 μl and placed on a rotating disk overnight at 4 °C, the resuspended DNA was stored at 4 °C rather than freezing at -20 °C. Some sequence data was obtained when using 1 μg purified cosmid, but the data had very high background and only a limited amount of sequence data was readable.

Cosmid DNA prepared using caesium chloride density gradient.

Cosmid DNA was prepared as previously described and varying amounts of the DNA (1, 3 and 6 μg) were used in sequencing reactions performed on a Perkin-Elmer Cetus 480 thermal cycler using the standard cycling parameters detailed in the Applied Biosystems DNA sequencing protocol (part no. 901497 Rev. C), purified using phenol / chloroform extraction (x3) and ethanol precipitation. Sequence data (250 bases) was obtained from 1 μg template, however, the ratio of "noise to signal" was high which resulted in some spurious peaks which in turn masked the real sequence data and required careful manual base calling (figure 25). The reactions were repeated using a Pharmacia.LKB Gene ATAQ and purified using Bio-Rad Biospin 30 columns but no readable sequence data was obtained.

Figure 25.

Direct sequencing of cosmid DNA.

Sequence data obtained from 1.0 μg cosmid DNA (clone 4) with primer 792.

Applied Biosystems

Model 373A

Version 1.2.0

011293_06

DyeTerminator{AnyPrimer}

Lane 6

Signal: G:245 A:259 T:260 C:62

Points 1331 to 5652 Base 1: 1331

MATRIX 077

JB: pAtG4.

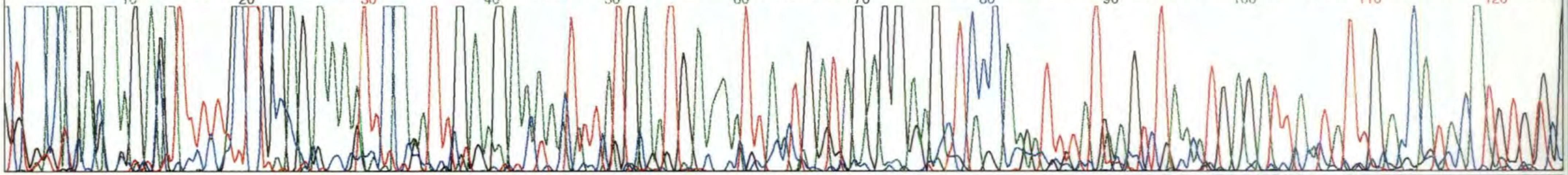
Primer 792.

Tue, Dec 1, 1992 12:15

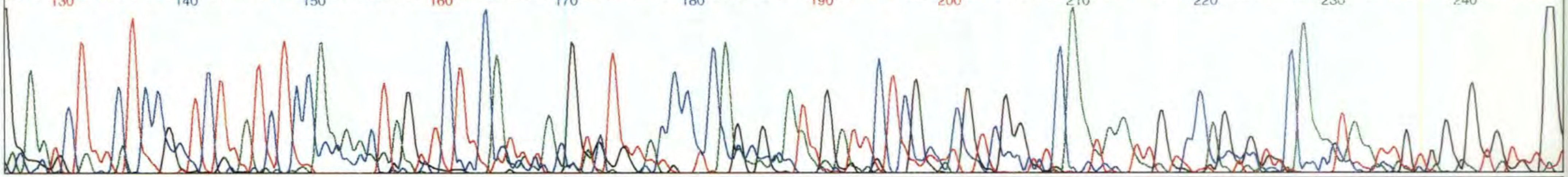
X: 0 to 4394 Y: 0 to 1200

Spacing: 12.34

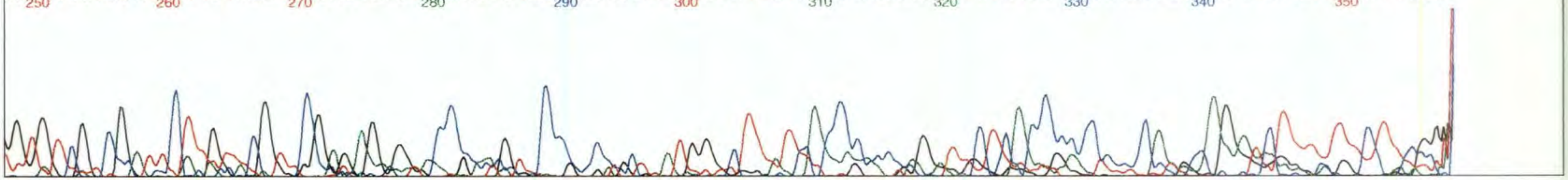
TCCACAGCAAGANATTTTNNTCGAGAAAATTCANATTGAAGAAAANTTTATGAATGAAAATTANTGATAGAGGAAGCTCCCAAATTTATNAGTTAAATGAGATTAATATTGACCATACATGTGNN



GCAATCTTTCTCCGCTCTTATCTCCAAAACTAGTTCTTCATNNACGTNTNTACCCTCAGCGNATTGATTCTCGCTCGTCGGTTCAATAATTGTCCAGNGTNCAAATAATTGTGGAGTGTNTG



NGTGTCGTCGATTCTTGTACGTTCGAGAGAGTACCNNAGTNCCGNCNGCTNATGGGCTTATCACCCACCCNGATTCTCACCCCCCCCCCANNCAGATCTNTTTTCTTCNGN



Purification of PCR products for DNA sequencing.

Purification of PCR products from agarose gels (Zhen and Swank, 1993).

This method has been used for either purifying PCR products prior to cloning into Promega pGem T vector or DNA sequencing. The yields (30-50 ng μl^{-1}) were not as high as obtained by Zhen and Swank, however, they were sufficient for cloning and more importantly direct DNA sequencing for obtaining rapid characterisation of the amplified products. When the DNA concentration had been quantified correctly an individual sequencing reaction may produce sequence data up to 450 bases in length.

Purification of PCR products from agarose gels using DEAE cellulose.

This protocol has been used for purifying PCR products prior to DNA sequencing. This method when followed correctly yielded very pure DNA for sequencing (100-200 ng μl^{-1}) and when the DNA concentration had been quantified correctly an individual sequencing reaction produced sequence data of 400-450 bases in length.

Commercially available PCR "clean-up" kits.

For relatively fast procedures for obtaining purified PCR products commercially available kits have been used. Greater than 99 % of DNA can be recovered when applying 50 ng-16 μg of a homogeneous 300 bp PCR product to 1 ml of Magic PCR Preps Resin. One drawback associated with these methods is when more than one PCR product is amplified within the reaction. This requires gel purification of the products using low melting point agarose prior to purification by one of the kits previously described. However, this resulted in a lower product yield than when purified directly. To obtain sequence data of high quality, an additional ethanol precipitation step was required when the products were purified by one of these kits. The products when used for sequencing are eluted in water, not TE or Tris buffer to minimise salt concentration, as too much salt will seriously inhibit *Taq* polymerase in the sequencing reaction.

DNA concentration for sequencing.

The DNA concentration is also very important, if the DNA concentration was estimated incorrectly, poor quality sequence data was obtained. A standard of known concentration should be loaded onto the agarose gel when quantifying samples for sequencing. Specific amounts of DNA in specific volumes are required depending on the sequencing chemistry used. When using cycle sequencing, the amount of DNA required is considerably less than for conventional sequencing.

Purification of primers for sequencing.

The performance of sequencing primers was very variable without any purification steps, therefore, an ethanol precipitation was introduced which greatly improved the sequence data and has now become standard procedure.

Summary of Research

Pea

PCR of $PsMT_A$ cDNA —› sequence of PCR product —› PCR of pea genomic DNA (figure 3) and probed with $PsMT_A$ cDNA PCR product (figure 4) —› 3 products cloned into pUC 18 and sequenced (figure 26) —› $PsMT_A$, $PsMT_B$ and $PsMT_C$ —› APCR (figures 5-9) —› screened 2 λ phage libraries with $PsMT_A$ cDNA PCR product —› hybridising plaques detected —› PCR of λ phage and probed with $PsMT_A$ cDNA PCR product —› multiple products detected (figure 10) —› isolation of λ phage DNA for use in PCR reaction —› all products $PsMT_A$ —› PCR of original libraries (figure 11) —› no products amplified.

Phaseolus vulgaris

PCR of genomic DNA and probed with $PsMT_A$ cDNA PCR product (figure 12).

Nicotiana tobacum

Isolation and characterisation of partial cDNA using heterologous probing with $PsMT_A$ and DNA sequencing —› screened λ phage library with *NtMT* partial cDNA —› 3 genomic clones detected by autoradiography (1, 2 and 13) —› nested deletions of the 3 genomic clones with subsequent DNA sequencing using the M13 forward primer —› primers designed for "gene walking" —› no Cys-rich motifs identified —› clone 13 showed amino acid sequence similarity to meri 5 sequence from A. thaliana (figure 28)

Arabidopsis thaliana

PCR of cDNA —› product cloned into pGEX 3X, sequenced and corresponded to *MT-2* cDNA (figure 30) —› screened cosmid library with *MT-2* cDNA PCR product —› 9 hybridising colonies detected —› restriction analysis (figure 14) —› direct sequencing of cosmid clones (2, 4, 6, 9) and clone 4 designated *AtMT-2* —› primer designed from cDNA to 'gene walk' into 5' flanking region using direct sequencing of cosmid DNA (figure 25) —› 276 bp of 5' flanking region obtained —› further restriction analyses —› 2 fragments cloned into pUC 18 —› recombinant DNA not identified.

Arabidopsis thaliana

PCR of cDNA —› product cloned into pGem T vector, sequenced and designated *MT-1* cDNA (figure 19 and 29) —› screened cosmid library with *MT-2* cDNA PCR product —› 2 hybridising colonies detected (A and B) (figure 18B) —› restriction analysis (figure 20) —› 3.5 kb *Pst* I fragment from A cloned into pUC 18 (figure 21) —› sequenced and designated *AtMT-1* —› primers designed from sequenced data to 'gene walk' into 5' flanking region (figure 29) —› 633 bp of 5' flanking region obtained.

CHAPTER 4

DISCUSSION

The polymerase chain reaction was used to amplify the coding regions of two additional metallothionein-like genes from pea. The partial sequencing of these genes, designated *PsMT$_B$* and *PsMT$_C$*, is described. The isolation and characterisation of a partial cDNA from *N. tabacum* (via heterologous probing) and its subsequent use to screen a *N. tabacum* genomic library is also described. Three *N. tabacum* genomic clones (1, 2 and 13) were identified, sequenced and were shown to have some sequence similarity to the partial cDNA sequence but further sequencing failed to identify any cysteine-rich motifs. However, clone 13 showed significant sequence similarity to a gene from *A. thaliana* designated *meri-5* .

The characterisation of the 5' flanking regions of two MT - like genes from *A. thaliana* is described. The genes, designated *AtMT-1* and *AtMT-2* with consideration to the location of the cysteine residues, were isolated by screening an *A. thaliana* genomic library with cDNA probes generated by PCR based upon known cDNA sequences (Takahashi, 1991; Zhou and Goldsbrough, 1993). The 5' flanking regions of *AtMT-1* and *AtMT-2* have been compared to each other and to those of the MT-like genes from pea (*PsMT$_A$*), maize (*MT-L*) and *N. crassa* (a class I MT) and searched for any conserved (putative regulatory) motifs. These comparisons were made with consideration to whether a particular gene had been isolated from a monocotyledenous or dicotyledenous species, designated a type 1, 2 or E$_C$ MT-like gene, or whether it represented one of a complement of known MT-like genes isolated from a particular species (from *A. thaliana*).

A PCR product has also been amplified using primers designed to a third MT - like cDNA sequence from *A. thaliana* (Raynal *et al* ., 1993). The sequence of this cDNA is known to have similarity, with respect to the locations of the encoded cysteine residues, to the E$_C$ gene from wheatgerm. These three genes not only differ with respect to the locations of the encoded cysteine residues but are also known to be expressed in different organs. Methodology for the optimisation of automated sequence analysis with respect to template preparation, direct sequencing of cosmid clones and PCR products is also discussed.

Isolation and Characterisation of homologues of *$PsMT_A$*.

Amplification of the coding regions of other members of the pea MT-like gene family.

The polymerase chain reaction and specific primers designed to *$PsMT_A$* (Evans *et al* ., 1990) were used to amplify the coding regions of two other members of the pea gene family, *$PsMT_B$* and *$PsMT_C$* (figure 3). The DNA sequences were compared to *$PsMT_A$* (figure 26) and some differences between the three sequences noted. Some of these differences (indicated in bold type) may have arisen during PCR and are of the normal A.T to G.C transitions (Saiki *et al* ., 1988). More clones would need to be sequenced and a consensus sequence obtained to establish if these differences were in fact errors introduced by *Taq* polymerase. Other non-transitional differences (indicated by under-lining) have been noted, however, only some of these differences change the amino acid sequences (indicated below the sequence of *$PsMT_C$*). The introns of these genes differ in length to *$PsMT_A$* (the intron for *$PsMT_B$* is shorter than that of *$PsMT_A$* whereas the intron of *$PsMT_C$* is longer) and the sequences of the introns are different indicating that the genes amplified by PCR genuinely correspond to different genes and are not merely PCR artifacts. The *PsMT* sequences are the same at the intron / exon junction (AA : GT - intron - AG : AT). The location of the intron - exon boundries for MT - like genes have been identified. For the MT - like genes from pea all the intron locations are at codon 17, for maize and *N. crassa* the locations are at codon 18, and for *A. thaliana* (*AtMT-1*) the location is at codon 19/20. The nucleotide sequences of the protein coding regions of the three *PsMT* genes are almost identical which suggests that these genes represent a sub-family within the proposed type 1 MT-like genes. PCR products corresponding to a putative fourth member of this pea gene

Figure 26.

Comparison of DNA sequences from $PsMT_A$, $PsMT_B$ and $PsMT_C$.

The sequence of primers 190 and 191 (5' and 3' respectively) are indicated by under-lining and the sequence of primers 284 and 283 (5' and 3' respectively) are indicated in bold type. Possible PCR errors of the A.T to G.C are indicated in bold type, the non-transitional differences are indicated by under-lining. It should be noted, that the base numbering refers to the $PsMT_A$ sequence, since the middle portion of the intron of $PsMT_C$ has yet to be characterised so the actual full length of this sequence is not known. The unknown portion of $PsMT_C$ is indicated using a dotted line.

		M S G C G C G S S C N C G D S C K	
$PsMT_A$	1	AT**GTCTGGATGTGGTTGTGG**AAGCAGTTGCAACTGTGGTGATAGCTGCAA	50
$PsMT_B$		GTCTGGATGTGGTTGTGGAAGCAGTTGCAACTGTGGGGAGAGCTGCAA	
$PsMT_C$		GTCTGGATGTGGTTGTGGAAGCAGTTGCAACTGTGGTGATAGCTGCAA	
$PsMT_A$		GTAAGGATCCACCACCTTAATTCTTTGTTGTTTTTCTGTATAATTTTTTC	100
$PsMT_B$		GTAAGGATCCACCACCTTAA**C**TCTTCCTTA....................	
$PsMT_C$		GTAAGGAT**T**CACCACCTTAATTCTTTCGAGTTTATTGTTTATCTTATGAA	
$PsMT_A$		ATTACAATTATTTGTATGTCTATTTTTAATCATATAGATGATTCTTTGGA	150
$PsMT_B$		....	
$PsMT_C$		TGAATCTTACATATATTTGTTAATATAATATTGTATTTTTTAAGTACTTT	
$PsMT_A$		GATTTTTTTAAATAATTTGTTTAGTTTTATCGCATCGAATAATATATGAT	200
$PsMT_B$		....	
$PsMT_C$		TTGGTTGGGTTTTATGAAATTAGATACTTCCATTTTTTATGGTTACTTTT	
$PsMT_A$		CTGAGCATGAGAAAAATAAATTTAATATAGACGGATTGTTTTTTATAAAT	250
$PsMT_B$		....	
$PsMT_C$		GACTATCCTCATTCTTATTATTATTCTAAAACAAAAATTAAAATAAAGTT	
$PsMT_A$		GAATTAGGCTGAATCTAAATTCTAAGACTATGAATATGGTTCATAATTCT	300
$PsMT_B$		....	
$PsMT_C$		CATATTTACTGAATAATTGATTAACCTGATTTTTATAATTCAGATTTTTT	
$PsMT_A$		ATGTTAAATCATTTTGTGTAGTGAAATTGGGCAATTTTATGTGTAAACGC	350
$PsMT_B$		....	
$PsMT_C$		AAATTTATTAAACAATTTAAATTTAAAATAAATTCATTTAAAAATTTAAA	
$PsMT_A$		ATAATTTTGAGGTTTAAAATAAGGATCGTGCTGTCGCGATAGTTTAAGTG	400
$PsMT_B$		....	
$PsMT_C$		AATCAGATTTATTGACATGAAATGTTTTAAGAAATTTGATTATTGGGATT	
$PsMT_C$		....	
$PsMT_A$		TCGATTGTAGTCGCGTAAAGGCTTTTCTGATTTCGGTTGGTTTAAGTGTT	450
$PsMT_B$		....	
$PsMT_C$			
$PsMT_A$		GATTGCAATCGTGTAAAGAGTTTTCTAATTTCGGTTGGTTTAGGTGCGAT	500
$PsMT_B$		....	
$PsMT_C$		....ATAG	
$PsMT_A$		TGCAGTCGGGTAAAGATTTTCGTGATTGTCGTCGTTGCGGTGTGAATTAA	550
$PsMT_B$		....	
$PsMT_C$		GCCAATTTTTCTCTTTCAGATTTTAACTGTCTAAAGACTATGAATTGATG	
$PsMT_A$		TCACAATTTCTTCTTTATCATAAAAACGTTGAATAACATATCGATATCGA	600
$PsMT_B$		CATTAAAATCTATATAATTATTTTTTTCTATATGGTTATTAATCTGTTGG	
$PsMT_C$		ATATTGAATCAATTGTCACTTGAATATGTTAAAAAGCTTTTTGATGGTAT	
$PsMT_A$		TTTGAAAACCTTTTTCGTGTAACGGTCTTTCGAAAACCTTTATTTTGACA	650
$PsMT_B$		TATTTATTTTTTTTCTTAAGAATCAATCTTGCAAAAAATTTGTTAATATT	
$PsMT_C$		TAAAATAAGATAAGGTTTCCCTTTAGCCGTCCATAAATTAAAAACATATA	

```
                                              C  N  K  R  S
PsMTA      ACCAAGTTTATAATTGATTTGTTTTGCTTGACAGATGCAACAAGAGGTCT   700
PsMTB      TTATTGGATTATTTTGATTTTTATTGCTGTGCAGATGCAACAAGAGGTCT
PsMTC      TTAATATGTATAATTGATTTTTATTGCTGTGCAGATGCAGCAAGAGGTCT
                                                 S

           S  G  L  S  Y  S  E  M  E  T  T  E  T  V  I  L  G
PsMTA      AGTGGATTGAGCTACTCCGAAATGGAAACCACCGAAACCGTGATTCTTGG   750
PsMTB      AGTGGATTGAGCTACTCCGAAATGGAAACCACAGAAACCGTGATTCTTGG
PsMTC      AGAGGATTGAGCTACTCCGAAATGGAAACCACAGAGACCGTGATTCTTGG

            V  G  P  A  K  I  Q  F  E  G  A  E  M  S  A  A  S
PsMTA      CGTCGGTCCGGCGAAGATCCAGTTTGAAGGTGCTGAAATGAGTGCTGCTT   800
PsMTB      CGTCGGTCCGGCGAAGATCCAGTTTGATGGTGCTGAAATGAGTGTTGCAG
PsMTC      CGTCGGTCCGGCGTTGATCCAGTTTGATGGTGCTGAAATGAGTGTTGCAG
                                    N                 V  A  A

             E  D  G  G  C  K  C  G  D  N  C  T  C  D  P  C
PsMTA      CTGAGGATGGTGGCTGCAAGTGTGGTGATAACTGCACTTGTGACCCTTGC   850
PsMTB      CTGAGGATGGTGGCTGCAAGTGTGGTGATAACTGCACCTGTGACCCCTGC
PsMTC      CTGAGGATGGTGGCTGCAAGTGTGGAGATAGCTGCACTTGTGACCCTTGC

           N  C  K
PsMTA      AACTGCAAATGAAGTGTAACATATAAAAGCTTGAAGCAGAGATATTGAAA   900
PsMTB      AACTGC............................................
PsMTC      AACTGCAAA.........................................

PsMTA      CCATTATGTTTAATTGTGTGTATGAGTACACATGTGTTTGGTTTTT (946)
```

family have been detected by autoradiography but sufficient product has not been visualised by agarose gel electrophoresis and ethidium bromide (EtBr) staining. An aliquot of the primary PCR was used as a template for a secondary PCR reaction. A band of ca. 1.2 kb was observed by EtBr staining, however, the amount of product was insufficient for purification and subsequent cloning.

Ligation-mediated PCR or anchored PCR (APCR).

As an alternative to screening genomic libraries for obtaining the 5' flanking region of individual genes, an attempt was made to use APCR to obtain the 5' flanking regions of $PsMT_B$ and $PsMT_C$. In order to test this approach, $PsMT_A$ was used since the 5' flanking region of this gene had already been characterised. The expected product of 488 bp, determined from the restriction map of $PsMT_A$ under these specific ligation conditions, was not detected by hybridisation to a specific $PsMT_A$ PCR probe (figures 6-9). However, equivalent APCR has been demonstrated in *Salmonella typhimurium* (Shyamala and Ames, 1989) and cyanobacteria (Robinson *et al*., 1990). A possible reason why this approach has been unsuccessful in this instance, is that in plants the genome size may be too great to obtain specific products based on a single specific primer, especially in pea which has a genome size of 5.0 x 10^9 bp, in comparison to the genome sizes of *Salmonella typhimurium* and cyanobacteria which are ca. 5.0 x 10^6 bp to 2.0 x 10^7 bp respectively. Alternatively, the use of nested PCR primers may improve the specificity of the reaction. A company, Takara Shuzo Co., Ltd. (Biomedical Group, Otsu, Shiga, Japan), supply PCR amplification and *in vitro* cloning kits for the amplification of unknown regions from cDNA and genomic DNA. The kit includes two sets of primers, one set to be used in a primary PCR reaction of which the products serve as template for a secondary PCR reaction using the second set of primers, which associate inside the region amplified by the first set of primers. This method should increase the amount of the specific product for cloning and sequencing purposes.

Screening pea genomic libraries.

Two pea genomic libraries were screened in an attempt to isolate the genes corresponding to $PsMT_B$ and $PsMT_C$. A number of plaques (15) from the two libraries hybridised to a $PsMT_A$ cDNA probe. Subsequent secondary and tertiary screening resulted in a number of plaques hybridising to the probe. PCR was performed on DNA prepared from plate lysates using primer set 3 (284 and 283). All the products observed were ca. 820 bp suggesting that all the hybridising plaques correspond to clones containing $PsMT_A$ (figure 10). A PCR reaction performed on the original phage libraries failed to amplify any products, but products corresponding to $PsMT_A$, $PsMT_B$ and $PsMT_C$ were amplified from the positive controls (figure 11). Since hybridising plaques were previously obtained from both libraries, phage containing inserts which corresponded to one or more of the *PsMT* genes should be present. Possible reasons why no products were observed may be due to incompatible cycling conditions, or that more template would be required, or a secondary PCR would be required to increase the amount of product amplified. Another approach for isolating the genes corresponding to $PsMT_B$ and $PsMT_C$ would be to design primers to both ends of the introns of $PsMT_B$ and $PsMT_C$, amplify both introns and use the products to rescreen the genomic libraries.

Amplification of a cognate from *Phaseolus vulgaris*

Initially, Sourthern analyses revealed cognates to $PsMT_A$ in genomic DNA from *P. vulgaris*. A PCR reaction was prepared and a fragment of ca. 700 bp was amplified from *P. vulgaris* genomic DNA (figure 12). Unfortunately, it has not been possible to clone this product into a suitable vector using T4 polymerase to "blunt" the ends of the PCR product. This work was done prior to the common use of TA vectors for cloning PCR products which could possibly facilitate cloning of such refractory products. Alternatively, heterologous probing of a *P. vulgaris* cDNA library might be used to search for putative cognates to

$PsMT_A$ which in turn would be used as probes to screen a genomic library and allow characterisation of such genes.

Isolation and characterisation of MT - like genes from *Nicotiana tabacum* .

A partial cDNA sequence isolated from *N. tabacum* (*NtMT* cDNA) which encoded Cys-Xaa-Cys motifs at the C-terminus (EFGEKAAEGGNGCKCGSNCTC DPCNC) was radiolabelled and used as a probe to screen a *N. tabacum* genomic library. Three clones (1, 2 and 13) showed extensive hybridisation to the probe. Further analysis of the clones using nested deletions and sequencing, revealed that clone 1 (no. 15) was homologous to the *NtMT* cDNA in a 280 base region, clone 2 (no. 5) was homologous in a 73 base region and clone 13 (no. 13) was homologous in a 120 base region (figure 13). Further sequence analysis by "gene walking" failed to identify any regions encoding cysteine-rich motifs. The sequence of clone 13 contained two ORFs. The first ORF (not full length) encoded 147 amino acids and the second ORF encoded 191 amino acids. The full sequence of clone 13 was compared the *NtMT* cDNA sequence (figure 27) and to sequences in the EMBL database and the amino acid sequence encoded by ORF 2 showed 84 % similarity to a *meri-5* sequence from *A. thaliana* (Medford *et al* ., 1991) (figure 28).

The above-ground portion of a plant develops from the shoot apical meristem and an abundant source of apical meristems is found in a *Brassica* mutant, commonly known as cauliflower. Meristematic cDNAs have been identified from cauliflower by differential screening and these used to isolate corresponding *A. thaliana* genomic clones (Medford *et al* ., 1991) One of the clones, designated *meri 5* , was characterised but no homologous sequences were found when the coding sequence was compared with the GenBank database (Medford *et al* ., 1991). The meristematic dome in the shoot apex, which is the source of development of the entire above-ground portion of a plant, functions to proliferate cells as well as to initiate new tissues and organs. Genes expressed in meristems may also be expressed in other tissues (Medford *et al*., 1991). To our knowledge the sequences of *meri 5* homologues have not previously been reported for other plant species including tobacco.

The sequence of the predicted protein encoded by ORF2 in clone 13 was significantly similar to the predicted product of *meri-5* (figure 28). It is unlikely that the 3' flanking regions of a tobacco homologue of *meri 5* and a MT-like gene would be so highly conserved. A possible explanation for the apparent sequence similarity between the two genes could be the occurrence of rearrangement during the construction of the cDNA library. It was observed that a poly A sequence (19 bases) was located 197 bases downstream of the *NtMT* cDNA (putative MT) coding region (figure 27) and could correspond to the polyA$^+$ tail of this transcript. It is possible that the 3' continuation of this cDNA sequence derived from a cloning artifact involving a fusion of the 3' end of the MT-like cDNA to a cDNA encoding a tobacco *meri 5* homologue. In order to isolate genomic clones corresponding to the *NtMT* cDNA, two primers may be designed to amplify a portion of the *NtMT* cDNA via PCR. One primer should be designed to the coding region, the other primer designed upstream of the poly A sequence and the resulting product could then be used to rescreen the genomic library.

Figure 27.

Comparison of clone 13 from *N. tabacum* genomic DNA sequence to an *NtMT* cDNA sequence.

The regions of homology are indicated by under-lining.

```
                 E  F  G  E  K  A  A  E  G  G  N  G  C  K  C  G  S
NtMT cDNA        GAATTCGGAGAGAAAGCAGCAGAAGGAGGAAATGGGTGCAAGTGTGGATC

                  N  C  T  C  D  P  C  N  C  *
NtMT cDNA        AAACTGCACCTGTGACCCTTGCAACTGTTAAGAATATTCTCACTACAAAA
Clone 13    1    CCCTGCTATATTACAACCCCACTAATGTTTCCAAATAAACAAAAGGTTTA     50

NtMT cDNA        ATAACAATAAAACTACAACCATGGAACAGATTATGATCAGTTCTTAGGTT
Clone 13         AATTCCAGCTGGTTTGGCAATTTTGGGTCCCTCCCCCCCAATTCATGAGC    100

NtMT cDNA        ATAACAATAAAACTACAACCATGGAACAGATTATGATCAGTTCTTAGGTT
Clone 13         ATGGGGAATCATATAATTGGCCCAATATAAATACATAGGCATTCTTGAAA    150

NtMT cDNA        GTGTGTTACTCGTGTGACGTGGCTTTTCTGTTTGTTAATTGGGTTGTGTA
Clone 13         AATTTCATATCATCAATCACCACACATTGATTTCAAACATCACCCCCCCC    200

NtMT cDNA        TCAATATACTACATAAGTTTGTTTAAAAAAAAAAAAAAAAAAAATTCTTGA
Clone 13         CCCCCCCCCCTTCTTCTTTAAACTTGCAAAAAAACTTTGAGAACATTTTT    250

NtMT cDNA        AATATTAGGTAGGAATTAAAGATGTCTTCTTTTTCTTCTAAATTAGTACT
Clone 13         GAGCTCTGAGAAATGTCGCCTCGTTTCTCTTTCAAAATGTTAATCCTTCC    300
                            [M  S  P  R  F  S  F  K  M  L  I  L  P

NtMT cDNA        AGCTCTTATTGTTAGTGCTTTCGCTATTGCAATTGCGGGTACTATTGACG
Clone 13         TATAGTCATGGCAAGTCTATGGGCAGTCGCCTCAGCTGGTAATTTTTATA    350
                  I  V  M  A  S  L  W  A  V  A  S  A  G  N  F  Y  N

NtMT cDNA        AAAATTTTGAAATTACATGGGGTGAAGGCAGAGCAAAGATGCTAAATAAT
Clone 13         ATCTTGCAGATATCACTTGGGGCGAAGGACGTGGTAAAATAACAGGAGGA    400
                   L  A  D  I  T  W  G  E  G  R  G  K  I  T  G  G

NtMT cDNA        GGAGAGCTTCTAACTCTATCACTTGACAAAATCTCAGGCTCAGGATTTCA
Clone 13         GGCAGAGGCCTTTCTCTGTCCCTTGATAAATTTTCTGGTTCGGGTTTTCA    450
                 G  R  G  L  S  L  S  L  D  K  F  S  G  S  G  F  Q

NtMT cDNA        ATCCAAGAATGAATATCTCTTTGGTAAAATAGACATGCAACTCAAACTTG
Clone 13         ATCGAAGAATGAGTATCTCTTTGGAAGATTTGACATGCAACTCAAACTTG    500
                  S  K  N  E  Y  L  F  G  R  F  D  M  Q  L  K  L  V

NtMT cDNA        TCCCTGGAAATTCTGCTGGCACTGTCACTGCTTACTATTTGTCATCACAA
Clone 13         TCCCTGGAAATTCTGCTGGCACTGTCACCACCTTCTTCGTAAGTCATTCA    550
                   P  G  N  S  A  G  T  V  T  T  F  F  V  S  H  S

NtMT cDNA        GGACCAACACATGATGAAATAGATTTTGAATT
Clone 13         ACTTTTCCCTTACAAAATTGTATTAGTTCTATTTCATGTCCAAGTTCTGA    600
                 T  F  P  L  Q  N  C  I  S  S  I  S  C  P  S  S  D

Clone 13         TGATTTAATATTTTTTCTTGTTCTTTGTTTCTTTCATCAGTTATCTTCAC    650
                  D  L  I  F  F  L  V  L  C  F  F  H  Q  L  S  S  Q
```

```
Clone 13      AAGGAGCTGGACATGATGAGATTGATTTCGAGTTCTTAGGCAATGTTTCT     700
                G   A   G   H   D   E   I   D   F   E   F   L   G   N   V   S

Clone 13      GGCCAACCTTACACAGTTCATACCAATGTTTACTCGCAAGGCAAAGGCAA     750
              G   Q   P   Y   T   V   H   T   N   V   Y   S   Q   G   K   G   N

Clone 13      CAAAGAACAACAATTCCATTTGTGGTTCGACCCAACTGCTGCATTTCACA     800
               K   E   Q   Q   F   H   L   W   F   D   P   T   A   A   F   H   T

Clone 13      CTTACTCCATTATCTGGAATGCTCAGAAGATCATGTAAGTTCCTTATATT     850
               Y   S   I   I   W   N   A   Q   K   I   M   *]

              AATCGCACTGTAAGAATTAATGATGTTTGCTTGCAATTTCTTCTTAGTTT     900
              GGTACTGATGATTTCCATATGAATTCCCATCTTCAGTTTCTTGGTAGATA
              ATAGTCCAATCAGAGTATACAACAACCACGAAAGCGCTGGCATTCCATTC
              CCAAAAAGCCAACCAATGAAAGTGTACTGCAGCTTATGGAATGCAGATGA
              GTGGGCTACACAAGGAGGTAGAGTCAAGACAGATTGGACACATGCTCCTT
              TCACTGCATATTACAGAAATTTCAATATTGATGCTGCGCAGTCACATCCG
              GCGCCTCTTCGTGTAAGTCCACTGATTCAGCAAACAATGCTAGGCCATGG
              CAAAATCAAGAACTTGATGCTAAGGGCAGGAATAGGCTACGATGGGTGCA
              GAGCAGACACATGGTTTACAACTATTGTGCTGATTCTAAGAGGTTTCCTC
              AAGGCTTTTCTCATGAGTGCAAGCGTTCGAGGTTCCTCGAGCTCGAATTC
              GCCCTATAGTGAGTCG  1366
```

Figure 28.

Amino acid sequence comparison of clone 13 from *N. tabacum* and Meri 5 from *A. Thaliana* .

```
                 1         .         .         .         .       50
Clone 13         MSPRFSFKMLILPIVMASLWAVASAGNFYNLADITWGEGRGKITGGGRGL
                 |||   ||::::..::.. :.| ||::| . .::.||:||||| ..|. |
Meri 5           MSP...FKIFFFTTLLVAAFSV.SAADFNTDVNVAWGNGRGKILNNGQLL
                 1                                                45

                 51        .         .         .         .      100
Clone 13         SLSLDKFSGSGFQSKNEYLFGRFDMQLKLVPGNSAGTVTTFFVSHSTFPL
                 .||||| ||||||||.|||||::|||:||||||||||||||:
Meri 5           TLSLDKSSGSGFQSKTEYLFGKIDMQIKLVPGNSAGTVTTFY........
                 46                                               88


                 101       .         .         .         .      150
Clone 13         QNCISSISCPSSDDLIFFLVLCFFHQLSSQGAGHDEIDFEFLGNVSGQPY
                                           |.|:|.. ||||||||||:||:||
Meri 5           ..........................LKSEGSTWDEIDFEFLGNMSGDPY
                 89                                              111


                 151       .         .         .       191
Clone 13         TVHTNVYSQGKGNKEQQFHLWFDPTAAFHTYSIIWNAQKIM........
                 |:|||||.||||:|||||||||||||.||||||:||:|:|:
Meri 5           TLHTNVYTQGKGDKEQQFHLWFDPTANFHTYSILWNPQRIILTVDTHPL
                 112                                             160
```

<u>Isolation and characterisation of MT - like genes from *Arabidopsis thaliana*.</u>

Two genes from *A . thaliana* have been isolated and characterised and their nucleotide sequences compared to the cDNA sequences used to isolate them (figures 29 and 30). *AtMT-1* contains an intron of 234 bp and there are sequence changes in three regions compared to that of the *MT-1* cDNA. The genomic clone is different in two positions at the region encoding N-terminus of the putative protein, bases 571 - 585 and an additional 17 bases at positions 634 - 650, and one position at the C-terminus where there is an 18 base deletion in the genomic clone at bases 990 - 1008. This deletion (CAG CTG TGG GTC AAA CTG TAA), except for one base which is indicated by underlining, is a direct repeat of the previous 18 bases, however, the amino acid sequence remains unchanged (S C G S N C) (figure 29). Other base changes are indicated in bold (figure 29). The nucleotide sequence of *AtMT - 2* was obtained by direct sequencing of the cosmid clones which hybridised to the cDNA probe. The PCR primer designed to the N-terminal of the corresponding cDNA (primer 519) was purified and used as a sequencing primer. Although the sequence contained a number of ambiguities, sequence data was determined for the first 43 bases and corresponded to the *MT-2* cDNA sequence. From this data, a primer was designed (primer 792) to "gene walk" into the 5' flanking region of the gene. Again the sequence contained some ambiguities but 276 bases 5' flanking region were determined using manual base-calling.

The disparity between the *AtMT-1* genomic sequence and the sequence of the cDNA probe could be due to the cDNA sequence hybridising to another member of a gene family. It is possible that the pattern of expression driven by the promoter of the *AtMT-1* sequence may be different to the original *MT-1* cDNA, or indeed it could be a pseudo gene which is not expressed at all. Alternatively, the differences could be attributed to possible sequencing errors or cloning artifacts. Comparison of the 5' flanking sequences of *AtMT-1* with other MT genes (refer to following section) will be valid if the latter two alternatives are correct, but may not be valid if the former are true. *AtMT-2* from *A. thaliana* showed some sequence similarity in the coding region of the original *MT-2* cDNA, however, to confirm that this clone corresponds to the cDNA, more sequence data will be required following sub-cloning. It is yet to be established if the third PCR product corresponds to the cDNA to which the primers were designed.

Analysis of the 5' flanking regions of MT-like genes from plants.

The upstream sequences of MT - like genes were searched for any elements which

Figure 29.

Comparison of an *AtMT-1* genomic DNA sequence to an *A. thaliana MT-1* cDNA sequence.

The sequence of primers 826 and 827 (5' and 3' respectively) are indicated by underlining. A putative TATA box is indicated by bold type and the sequencing primers 894 and 902 are indicated by bold type and underlining.

```
AtMT-1    1 ACCCAATGGATTGATGATGTAATCCATGGATTGATGGCGGCACTTAGTCG   50
            TCCCAAAACTTTTTTTTTTTGTGTCAACAATTTTTAAAAAAATCTTAAAAC
            GGGATATCTGATGTAATATATGGTTCGTAAGAAACTCATATATAATATTA
            ATAAGTAGTAGAAGAGCGTAGACCATGTCCTGGGTCGTCGTCCAAATGAC
            CACAAGAAGATTTCAAAACAGAGGAAAATATTTCTCATTAAATAAGTTTT
            CCTGCCTCTTTTTTGGGTCGGCTGCCGTTTACTATTGGCAATTGCGATAG
            TTATCTTCTTGCGTGTACGACTTTTTTTGTTTTTTACATATTCGCCAATA
            ATTTGACGTTTTCTATTAGTTTGTTTGATACTCTGTTGTCTTGCTAAAAC
            TCTAACATTAAATTACTTTCTTGAATGAAGCTGGAACAAATCTAACATAA
            AAAGAAAATGATGGGCAAGTTGATGTTATTCGTAAATTTATTTGGATTAT
            GTATAAAGTGATCCGAATCTCTCTCTTCTTCTATAAATAGTGGCCATTCC

AtMT-1      CATTGGTTGAAATCACAAAGCATCATAAGAAGAAGAAGAAACTACAGAAG   600
  MT-1      ....................GAATTCGGCACGAGGAAGAAACTACAATAG

                                        M  G              D
AtMT-1      TTAATCAATCAAAGACAAGTAAGAGAAATGGGACTGATTCTAACCAGGAT   650
  MT-1      TTAATCAATCAAAGAGAAGTAAGAGAAATGGCA.................
                                        M  A

            D  S  N  C  G  C  G  S  S  C  K  C  G  D  S  C  R
AtMT-1      GATTCTAATTGTGGATGTGGCTCCTCCTGCAAATGTGGTGACTCTTGCAG   700
  MT-1      GATTCTAACTGTGGATGTGGCTCCTCCTGCAAATGTGGTGACTCTTGCAG
            D  S  N  C  G  C  G  S  S  C  K  C  G  D  S  C  R

AtMT-1      GTAACCCCCTTGATCCTCTCTCTCGTCTTTACCTTTAATGGCTTCATGGT   750
  MT-1      ..................................................

AtMT-1      ACAAAATTCATAAACTCTTCTTATGACTTACGTTTACAATTGTGTTAGTG   800
  MT-1      ..................................................

AtMT-1      AGTAGACACAAAAACATTTTGGCCTGTTCTGCATATATGCTGATCAGAAT   850
  MT-1      ..................................................

AtMT-1      CAGATGCATGTGGTTGACTAATCTTTAATCTCATCATATTAATTACTAAT   900
  MT-1      ..................................................

                                                C  E  K  N  Y
AtMT-1      TTTTTTGTTTGTTTGATTAATTTGTGTGTGGTATAGTTGCGAGAAGAACT   950
  MT-1      ....................................TTGCGAGAAGAACT
                                                C  E  K  N  Y

             N  K  E  C  D  N  C  S  C  G  S  N  C
AtMT-1      ACAACAAGGAGTGCGACAACTGTAGCTGTGGGTCAAACTG..........   1000
  MT-1      ACAACAAGGAGTGCGACAACTGTAGCTGTGGATCAAACTGCAGCTGTGGG
             N  K  E  C  D  N  C  S  C  G  S  N  C  S  C  G

                      N  F  *
AtMT-1      ........TAACTTCTGATGAAATTACTATGGTCTAAAATCATATATATG   1050
  MT-1      TCAAACTGTAACTGTTGATGAAATTATTATGGTCTAAAATCATATATATG
            S  N  C  N  C  *

AtMT-1      GCAGAAAAATTGGGGAAAATATGTGTTTTATGCTAAGAGATGTGTGTG..
  MT-1      GCAGAAAAATTGGGGAAAATATGTGTTTTATGCTAAGAGATGTGTGTGTG

  MT-1      TTGTTGGAATAAAGACGTGACCGTTGTGTTGCGTATCAACTCTCTTAAGC
            TTTGACTTTTCCCAGCTTTGTATTTTCCTATGTATGGTAATGGTGTGATT
            GTGTAATGTTTTCATATGTAACGTAAAAAAAAATATTTATGTGACATTGAC
            TTTTGTGACTACTAAAAAAAAAAAAAAAAAAA 1282
```

Figure 30.

Comparison of an *AtMT-2* genomic DNA sequence to an *A. thaliana MT-2* cDNA sequence.

The sequence of primers 519 and 518 (5' and 3' respectively) are indicated by underlining. A putative TATA box is indicated by bold type and the sequencing primer 792 is indicated by bold type and underlining.

```
AtMT-2    1  CGGAACATCGGTACGCTCTCGAACGTACAAGAATCGACGACACACAAACA   50
AtMT-2       CTCCACAATTATTTGAACACTGGACAATTATTGAACCGACGTACGAGAAT
AtMT-2       CAATGCGCTGAGGGTAAAGACGTAAATGAAGAACTAGTTTTGGAGATAAG
AtMT-2       AGCGGAGAAAGATTGCGACACATGTATGGTCAATATTAATCTCATTTAAC

AtMT-2       TTATAAATTTGGGAGCTTCCTCTATCATTAATTTTCATTCATAAATTTTT   250
  MT-2       ...................................CATTCATAAATTTTT

                                         M  S  C  C  G  G  N  C
AtMT-2       CTTCAATTTGAATTTTCTCGAGAAAAATGTCTTGCTGTGGAGGAAACTGC   300
  MT-2       CTTCAATTTGAATTTTCTCGAGAAAAATGTCTTGCTGTGGAGGAAACTGC
                                         M  S  C  C  G  G  N  C

AtMT-2       GG 302
  MT-2       GGATGTGGATCTGGCTGCAAGTGCGGCAACGGTTGTGGAGGTTGCAAAAT
             G  C  G  S  G  C  K  C  G  N  G  C  G  G  C  K  M

AtMT-2       ..................................................
  MT-2       GTACCCTGACTTGGGATTCTCCGGCGAGACAACCACAACTGAGACTTTTG
              Y  P  D  L  G  F  S  G  E  T  T  T  T  E  T  F  V

AtMT-2       ..................................................
  MT-2       TCTTGGGCGTTGCACCGGCGATGAAGAATCAGTACGAGGCTTCAGGGGAG
               L  G  V  A  P  A  M  K  N  Q  Y  E  A  S  G  E

AtMT-2       ..................................................
  MT-2       AGTAACAACGCTGAGAGCGATGCTTGCAAGTGTGGATCTGACTGCAAGTG
             S  N  N  A  E  S  D  A  C  K  C  G  D  N  C  T  C

AtMT-2       ..................................................
  MT-2       TGATCCTTGCACCTGCAAGTGAAGAAGCCTTTTTAAATAAGCAGAGATAA
              N  P  C  N  C  G  K

  MT-2       TCGAGTCTCTTTAATNTAATTAAGTTATTCAATAAGTAAACCATATATAG
             GATGGTGTTTTTAGGTTTGGTTTATGTGTAATAATGGCTTCAGCTTATCT
             TTTAGCCGATCATTGTCTTTTGTGTTTGTTTTGATCATATCTTTTAGATG
             TCTAGCAAATCTGCCATGTGATGAGTTTGTACTTCCAGTGGAATGATAAT
             AATATTATAGTTTTAAATCAAAAAAAAAAA
```

may be involved in the regulation of expression of plant MT - like genes. In animals the core MRE, which is conserved in MT genes, has been identified to be 5' TGCRCNCX 3' (where R represents G or A, X represents G or C and N can be any base but A) (Stuart *et al.*, 1985; Searle *et al* ., 1985). There is no reason to believe that expression of plant MT - like genes will involve similar elements, although a somewhat related sequence (5' TGCACACC 3') was found in the 5' flanking region of $PsMT_A$ (Evans *et al* ., 1990) (figure 31). In order to test the predicted regulatory function of the MRE sequences in the mouse MT-I gene, Searle *et al* ., (1987) systematically introduced single nucleotide changes. Some of these changes were found to abolish function within the core sequence of the MRE but certain changes outside this sequence were shown to have lesser effects. The related sequence identified from the $PsMT_A$ gene was found to have an A nucleotide at the sixth position of the core element instead of the stated C/G/T nucleotides. When an A nucleotide was introduced into the core element in the experiments stated above, a high level of basal expression was observed (in either the presence or absence of zinc).

Previous to this study, there were only two reports which presented 5' flanking sequences of MT - like genes from plants. One sequence was from maize, a monocotyledenous species, and a second from pea which is a dicotyledenous species. These two sequences have been categorised as type 1 MT - like genes with respect to the locations of the encoded cysteine residues. The 5' flanking regions of a type 2 MT - like plant gene was not previously known. From this study, two genes from the dicotyledenous species *A. thaliana, AtMT-1* and *AtMT-2* , have been partly characterised (figures 29 and 30) .

Various comparisons were made to identify any direct or indirect repeats within a particular sequence, and to identify any consensus sequences between the individual genes. This included the use of matrix comparisons to identify possible conserved motifs. These are summarised in table 2 and located as figure 31A-J to reveal the distance from the putative TATA box of each gene.

On comparison of the genes designated *PsMT*$_A$ from pea and *AtMT-1* from *A. thaliana* the motif TCGCCA $_n$ AATTTG (where n = 2 to 4 nucleotides) was identified (figure 31A). This may be significant as this motif has been identified in the 5' flanking regions of two plant sequences which are categorised as type 1 MT - like genes, are both from dicotyledenous species and are thought to be expressed in roots in response to copper ions (while noting the reservations about the identity of *AtMT-1* stated above). These motifs are both located at similar distances from the putative TATA boxes of the two genes. This provides some support to the suggestion that these may be functional motifs . Similar motifs have not been identified within any of the other flanking sequences. Two short motifs at similar postions in both species are also noted (table 2). It will be of interest to establish whether or not similar sequences occur in the flanking sequences of a third type 1 sequence. This awaits the isolation of an additional clone from a third dicotyledenous plant. The motif TCTCTCd TATA box bGGC has also been identified in both species and a similar motif present in the same region from *AtMT-2* (TCTCeTATAaGGG). These motifs, summarised in table 2, were not identified in the other species, and it remains to be established whether or not this is a common motif to all MT-like genes from dicotelydenous species.

There were a number of conserved motifs identified at similar positions from the TATA box when comparing *PsMT*$_A$ with *AtMT-2* (table 2). Similar motifs were not identified in the other 5' flanking sequences. A "somewhat" similar sequence to that of the conserved MRE from animal MTs was identified in the *AtMT-2* 5' flanking region. However, the presence of a G (underlined) rather than C within this sequence 5' TGCGCT<u>G</u>A 3' (indicated in bold type figure 31B) is known to abolish its function as an MRE in animals (Searle *et al* ., 1987) and, of course, there is no evidence that equivalent motifs would be functional in plants.

Comparison of the sequences from pea (*PsMT*$_A$) and maize (*MT-L*) revealed no significant conserved sequences which might be expected due to the diversity of the two species (dicotelydenous and monocotyledenous species respectively) (figure 31C).

Comparison of the 5' flanking sequences from pea (*PsMT$_A$*) and *Neurospora crassa* (*CuMT*) identified some conserved sequences reported by Evans *et al*., (1990) (the sequence from *N. crassa* is not indicated on figure 31D since it occurs 429 bases 5' of the TATA box). No other related sequences were reported when comparing *PsMT$_A$* and *CuMT* with the three remaining sequences (*MT-L*, *AtMT-1* and *AtMT-2*) (figure 31A-C).

A consensus motif was observed in the region of the putative TATA boxes of *AtMT-1* and *AtMT-2* TCTCaTTb TATA box bGGbTTCC (figure 31E), however, it is yet to be established if this motif is specific to the MT - like genes from *A. thaliana* or whether it is a conserved sequence present surrounding a number of *A. thaliana* promotor sequences, although it is certainly not present around all *A. thaliana* TATA boxes (eg absent from *meri 5*). Two short motifs were also identified at similar positions upstream of the TATA box (table 2).

The sequence comparison of *AtMT-2* from *A. thaliana* and *MT-L* from maize, identified four short sequences at similar positions from the TATA box (table 2). The possibility that these may be functional motifs seems unlikely based on the diversity of the two species.

On searching the sequences for direct and inverse repeats, a direct repeat AACACTaACAATTATTaAACACTaACAATTATT (where a = 2) was identified from the 5' flanking region of *AtMT-2*. At present it is not apparent if this is a conserved motif in type 2 MT-like genes from dicotyledenous.

From these analyses, the characterisation of the 5' flanking region of a third type-1 MT-like gene from a dicotyledenous species may confirm that the motif identified from *PsMT$_A$* and *AtMT-1* is a conserved motif. Also characterisation of a second type 1 MT-like gene from a monocotyledenous species and a second type 2 MT-like gene from a dicotyledenous species would possibly identify other conserved motifs within a given type of MT-like genes.

Figure 31.

Comparison of DNA sequences *PsMT$_A$* , *AtMT-1* , *AtMT-2* , *MT-L* , and *N. crassa* using the putative TATA boxes for alignment.

Putative TATA boxes and ATG translation start codon are indicated in bold type. Any similar sequences between two sequences are indicated by underlining.

A. *PsMT* x *AtMT-1*

PsMT	ACCTACATTTGACTTAGATTGGATTAAGCATGCAACAAATTAAAATTTAA	-151
AtMT-1	TTTTTACATATTCGCCAATAATTTGACGTTTTCTATTAGTTTGTTTGATA	
PsMT	TCGCCATTGCAATT**TGCACACC**ACAATAAGACGTGTGATGAAAACGATGA	-101
AtMT-1	CTCTGTTGTCTTGCTAAAACTCTAACATTAAATTACTTTCTTGAATGAAG	
PsMT	TATCTACGTGGAAATAATCCAAGGGTGGCCTTGTGGACCCATGCAACACA	-51
AtMT-1	CTGGAACAAATCTAACATAAAAAGAAAATGATGGGCAAGTTGATGTTATT	
PsMT	GGATGACAACACGTGGACGGTCAAGATTTCACCAATTATTCTCTCCCACC	-1
AtMT-1	CGTAAATTTATTTGGATTATGTATAAAGTGATCCGAATCTCTCTCTTCTT	
	+1	
PsMT	**TTATAAAT**TGGGGCACGCAACATCATTAAAAGACATCAATTGTAGTGAAG	50
AtMT-1	**CTATAAAT**AGTGGCCATTCCCATTGGTTGAAATCACAAAGCATCATAAGA	
PsMT	ATAACAGCAACCAAGCAATTAATATCAATTGTTGTTTGCAAAAAATCTTA	100
AtMT-1	AGAAGAAGAAACTACAGAAGTTAATCAATCAAAGACAAGTAAGAGAA**ATG**	
PsMT	GGTTCTGAAAAT**ATG**TCTGGATGTGGTTGTGGAAGCAGTTGCAACTGTGG	150
AtMT-1	GGACTGATTCTAACCAGGATGATTCTAATTGTGGATGTGGCTATTGTGGA	

B. *PsMT* x *AtMT-2*

PsMT	ACCTACATTTGACTTAGATTGGATTAAGCATGCAACAAATTAAAATTTAA	-151
AtMT-2	CGGAACATCGGTACGCTCTCGAACGTACAAGAATCGACGACACACAAACA	
PsMT	TCGCCATTGCAATT**TGCACACC**ACAATAAGACGTGTGATGAAAACGATGA	-101
AtMT-2	CTCCACAATTATTTGAACACTGGACAATTATTGAACCGACGTACGAGAAT	
PsMT	TATCTACGTGGAAATAATCCAAGGGTGGCCTTGTGGACCCATGCAACACA	-51
AtMT-2	CAA**TGCGCTGA**GGGTAAAGACGTAAATGAAGAACTAGTTTTGGAGATAAG	
PsMT	GGATGACAACACGTGGACGGTCAAGATTTCACCAATTATTCTCTCCCACC	-1
AtMT-2	AGCGGAGAAAGATTGCGACACATGTATGGTCAATATTAATCTCATTTAAC	
	+1	
PsMT	**TTATAAAT**TGGGGCACGCAACATCATTAAAAGACATCAATTGTAGTGAAG	50
AtMT-2	**TTATAAAT**TTGGGAGCTTCCTCTATCATTAATTTTCATTCATAAATTTTT	
PsMT	ATAACAGCAACCAAGCAATTAATATCAATTGTTGTTTGCAAAAAATCTTA	100
AtMT-2	CTTCAATTTGAATTTTCTCGAGAAAA**ATG**TCTTGCTGTGGAGGAAACTGC	
PsMT	GGTTCTGAAAAT**ATG**TCTGGATGTGGTTGTGGAAGCAGTTGCAACTGTGG	150
AtMT-2	GG 102	

C. *PsMT* x *MT-L*

```
PsMT        ACCTACATTTGACTTAGATTGGATTAAGCATGCAACAAATTAAAATTTAA   -151
MT-L        TTACAAATAAAGAATAAAGCTAGGACGAACTCGTGGATTATTACTAAATC

PsMT        TCGCCATTGCAATTTGCACACCACAATAAGACGTGTGATGAAAACGATGA   -101
MT-L        GAAATGGACGTAATATTCCAGGCAAGAATAATTGTTCGATCAGGAGACAA

PsMT        TATCTACGTGGAAATAATCCAAGGGTGGCCTTGTGGACCCATGCAACACA    -51
MT-L        GTGGGGCATTGGACCGGTTCTTGCAAGCAAGAGCCTATGGCGTGGTGACA

PsMT        GGATGACAACACGTGGACGGTCAAGATTTCACCAATTATTCTCTCCCACC     -1
MT-L        CGGCGCGTTGCCCATACATCATGCCTCCATCGATGATCCATCCTCACTTG

           +1
PsMT        TTATAAATTGGGGCACGCAACATCATTAAAAGACATCAATTGTAGTGAAG     50
MT-L        CTATAAAAAGAGGTGTCCATGGTGCTCAAGCTCAGCCAAGCAAATAAGAC

PsMT        ATAACAGCAACCAAGCAATTAATATCAATTGTTGTTTGCAAAAAATCTTA    100
MT-L        GACTTGTTTCATTGATTCTTCAAGAGATCGAGCTTCTTTTGCACCACAAG

PsMT        GGTTCTGAAAATATGTCTGGATGTGGTTGTGGAAGCAGTTGCAACTGTGG    150
MT-L        GTCGAGGATGTCTTGCAGCTGCGGATCAAGCTGCGGCTGCGGCTCAAGCT
```

D. *PsMT* x *N.crassa*

```
PsMT        ACCTACATTTGACTTAGATTGGATTAAGCATGCAACAAATTAAAATTTAA   -151
N.crassa    ACGGCACAAATATCGCAAACCGTGCGGGCGGACGTCCCCGATAGGCTCGG

PsMT        TCGCCATTGCAATTTGCACACCACAATAAGACGTGTGATGAAAACGATGA   -101
N.crassa    ATCGTCGCTAACATGCCATGCCATGCCATGCCATGCCGTGTCGTGGAATG

PsMT        TATCTACGTGGAAATAATCCAAGGGTGGCCTTGTGGACCCATGCAACACA    -51
N.crassa    GAAGGTAACAGACAGTCACAGTCACAGTTCCGAACGGTGCGTAGCAGTGT

PsMT        GGATGACAACACGTGGACGGTCAAGATTTCACCAATTATTCTCTCCCACC     -1
N.crassa    AGTGTACGTGACAGCAGGAACAAGGTACCGTACGTAGGCATCATTCAGGG

           +1
PsMT        TTATAAATTGGGGCACGCAACATCATTAAAAGACATCAATTGTAGTGAAG     50
N.crassa    TATATAAAGCCACGGGACCCTGACTCCTCCTCAAATTCCTTTACTTGTCA

PsMT        ATAACAGCAACCAAGCAATTAATATCAATTGTTGTTTGCAAAAAATCTTA    100
N.crassa    TCAACCGACGAACTAAATCATCACTGCATTGAAACACTCACACAAAGACA

PsMT        GGTTCTGAAAATATGTCTGGATGTGGTTGTGGAAGCAGTTGCAACTGTGG    150
N.crassa    ACATCAGTCTTTCGTCAACCTACACAACCACAACAACAACTCAATTCCTC

PsMT        TGATAGCTGCAAGTAAGGATCCACCACCTTAATTCTTTGTTGTTTTTCTG    200
N.crassa    TCAAAAGAACAACCAGTTCTATCAAAATGGGTGACTGCGGCTGCTCCGGC
```

E. *AtMT-1* x *AtMT-2*

AtMT-1	TTTTTACATATTCGCCAATAATTTGACGTTTTCTATTAGTTTGTTTGATA	-151
AtMT-2	CGGAACATCGGTACGCTCTCGAACGTACAAGAATCGACGACACACA**AACA**	
AtMT-1	CTCTGTTGTCTTGCTAAAACTCTAACATTAAATTACTTTCTTGAATGAAG	-101
AtMT-2	**CT**CC**ACAATTATTT**G**AACACT**GG**ACAATTATT**GAACCGACGTACGAGAAT	
AtMT-1	CTGGAACAAATCTAACATAAAAAGAAAATGATGGGCAAGTTGATGTTATT	-51
AtMT-2	CAATGCGCTGAGGGTAAAGACGTAAATGAAGAACTAGTTTTGGAGATAAG	
AtMT-1	CGTAAATTTATTTGGATTATGTATAAAGTGATCCGAATCTCTCTCTTCTT	-1
AtMT-2	AGCGGAGAAAGATTGCGACACATGTATGGTCAATATTAATCTCATTTAAC	
	+1	
AtMT-1	**CTATAAAT**AGTGGCCATTCCCATTGGTTGAAATCACAAAGCATCATAAGA	50
AtMT-2	**TTATAAAT**TTGGGAGCTTCCTCTATCATTAATTTTCATTCATAAATTTTT	
AtMT-1	AGAAGAAGAAACTACAGAAGTTAATCAATCAAAGACAAGTAAGAGAA**ATG**	100
AtMT-2	CTTCAATTTGAATTTTCTCGAGAAAA**ATG**TCTTGCTGTGGAGGAAACTGC	

F. *AtMT-1* x *MT-L*

AtMT-1	TTTTTACATATTCGCCAATAATTTGACGTTTTCTATTAGTTTGTTTGATA	-151
MT-L	TTACAAATAAAGAATAAAGCTAGGACGAACTCGTGGATTATTACTAAATC	
AtMT-1	CTCTGTTGTCTTGCTAAAACTCTAACATTAAATTACTTTCTTGAATGAAG	-101
MT-L	GAAATGGACGTAATATTCCAGGCAAGAATAATTGTTCGATCAGGAGACAA	
AtMT-1	CTGGAACAAATCTAACATAAAAAGAAAATGATGGGCAAGTTGATGTTATT	-51
MT-L	GTGGGGCATTGGACCGGTTCTTGCAAGCAAGAGCCTATGGCGTGGTGACA	
AtMT-1	CGTAAATTTATTTGGATTATGTATAAAGTGATCCGAATCTCTCTCTTCTT	-1
MT-L	CGGCGCGTTGCCCATACATCATGCCTCCATCGATGATCCATCCTCACTTG	
	+1	
AtMT-1	**CTATAAAT**AGTGGCCATTCCCATTGGTTGAAATCACAAAGCATCATAAGA	50
MT-L	**CTATAAAA**AGAGGTGTCCATGGTGCTCAAGCTCAGCCAAGCAAATAAGAC	
AtMT-1	AGAAGAAGAAACTACAGAAGTTAATCAATCAAAGACAAGTAAGAGAA**ATG**	100
MT-L	GACTTGTTTCATTGATTCTTCAAGAGATCGAGCTTCTTTTGCACCACAAG	
AtMT-1	GGACTGATTCTAACCAGGATGATTCTAATTGTGGATGTGGCTCCTCCTGC	150
MT-L	GTCGAGG**ATG**TCTTGCAGCTGCGGATCAAGCTGCGGCTGCGGCTCAAGCT	

G. *AtMT-1* x *N.crassa*

AtMT-1	TTTTTACATATTCGCCAATAATTTGACGTTTTCTATTAGTTTGTTTGATA	-151
N.crassa	ACGGCACAAATATCGCAAACCGTGCGGGCGGACGTCCCCGATAGGCTCGG	
AtMT-1	CTCTGTTGTCTTGCTAAAACTCTAACATAAATTACTTTCTTGAATGAAGT	-101
N.crassa	ATCGTCGCTAACATGCCATGCCATGCCATGCCATGCCGTGTCGTGGAATG	
AtMT-1	CTGGAACAAATCTAACATAAAAAGAAAATGATGGGCAAGTTGATGTTATT	-51
N.crassa	GAAGGTAACAGACAGTCACAGTCACAGTTCCGAACGGTGCGTAGCAGTGT	
AtMT-1	CGTAAATTTATTTGGATTATGTATAAAGTGATCCGAATCTCTCTCTTCTT	-1
N.crassa	AGTGTACGTGACAGCAGGAACAAGGTACCGTACGTAGGCATCATTCAGGG	
	+1	
AtMT-1	**CTATAAAT**AGTGGCCATTCCCATTGGTTGAAATCACAAAGCATCATAAGA	50
N.crassa	**TATATAAA**GCCACGGGACCCTGACTCCTCCTCAAATTCCTTTACTTGTCA	
AtMT-1	AGAAGAAGAAACTACAGAAGTTAATCAATCAAAGACAAGTAAGAGAA**ATG**	100
N.crassa	TCAACCGACGAACTAAATCATCACTGCATTGAAACACTCACACAAAGACA	
AtMT-1	GGACTGATTCTAACCAGGATGATTCTAATTGTGGATGTGGCTCCTCCTGC	150
N.crassa	ACATCAGTCTTTCGTCAACCTACACAACCACAACAACAACTCAATTCCTC	
AtMT-1	AAATGTGGTGACTCTTGCAGGTAACCCCCTTGATCCTCTCTCTCGTCTTT	200
N.crassa	TCAAAAGAACAACCAGTTCTATCAAA**ATG**GGTGACTGCGGCTGCTCCGGC	

H. *AtMT-2* x *MT-L*

AtMT-2	CGGAACATCGGTACGCTCTCGAACGTACAAGAATCGACGACACACAAACA	-151
MT-L	TTACAAATAAAGAATAAAGCTAGGACGAACTCGTGGATTATTACTAAATC	
AtMT-2	CTCCACAATTATTTGAACACTGGACAATTATTGAACCGACGTACGAGAAT	-101
MT-L	GAAATGGACGTAATATTCCAGGCAAGAATAATTGTTCGATCAGGAGACAA	
AtMT-2	CAATGCGCTGAGGGTAAAGACGTAAATGAAGAACTAGTTTTGGAGATAAG	-51
MT-L	GTGGGGCATTGGACCGGTTCTTGCAAGCAAGAGCCTATGGCGTGGTGACA	
AtMT-2	AGCGGAGAAAGATTGCGACACATGTATGGTCAATATTAATCTCATTTAAC	-1
MT-L	CGGCGCGTTGCCCATACATCATGCCTCCATCGATGATCCATCCTCACTTG	
	+1	
AtMT-2	**TTATAAAT**TTGGGAGCTTCCTCTATCATTAATTTTCATTCATAAATTTTT	50
MT-L	**CTATAAAA**AGAGGTGTCCATGGTGCTCAAGCTCAGCCAAGCAAATAAGAC	
AtMT-2	CTTCAATTTGAATTTTCTCGAGAAAA**ATG**TCTTGCTGTGGAGGAAACTGC	100
MT-L	GACTTGTTTCATTGATTCTTCAAGAGATCGAGCTTCTTTTGCACCACAAG	
AtMT-2	GG 102	
MT-L	GTCGAGG**ATG**TCTTGCAGCTGCGGATCAAGCTGCGGCTGCGGCTCAAGCT	150

I. *AtMT-2 x N.crassa*

AtMT-2	CGGAACATCGGTACGCTCTCGAACGTACAAGAATCGACGACACACAAACA	-151
N.crassa	ACGGCACAAATATCGCAAACCGTGCGGGCGGACGTCCCCGATAGGCTCGG	
AtMT-2	CTCCACAATTATTTGAACACTGGACAATTATTGAACCGACGTACGAGAAT	-101
N.crassa	ATCGTCGCTAACATGCCATGCCATGCCATGCCATGCCGTGTCGTGGAATG	
AtMT-2	CAATGCGCTGAGGGTAAAGACGTAAATGAAGAACTAGTTTTGGAGATAAG	-51
N.crassa	GAAGGTAACAGACAGTCACAGTCACAGTTCCGAACGGTGCGTAGCAGTGT	
AtMT-2	AGCGGAGAAAGATTGCGACACATGTATGGTCAATATTAATCTCATTTAAC	-1
N.crassa	AGTGTACGTGACAGCAGGAACAAGGTACCGTACGTAGGCATCATTCAGGG	
	+1	
AtMT-2	**TTATAAAT**TTGGGAGCTTCCTCTATCATTAATTTCATTCATAAATTTTT	50
N.crassa	**TATATAAA**GCCACGGGACCCTGACTCCTCCTCAAATTCCTTTACTTGTCA	
AtMT-2	CTTCAATTTGAATTTTCTCGAGAAAA**ATG**TCTTGCTGTGGAGGAAACTGC	100
N.crassa	TCAACCGACGAACTAAATCATCACTGCATTGAAACACTCACACAAAGACA	
AtMT-2	GG 102	
N.crassa	ACATCAGTCTTTCGTCAACCTACACAACCACAACAACAACTCAATTCCTC	150
N.crassa	TCAAAAGAACAACCAGTTCTATCAAA**ATG**GGTGACTGCGGCTGCTCCGGC	200

J. *MT-L x N.crassa*

MT-L	TTACAAATAAAGAATAAAGCTAGGACGAACTCGTGGATTATTACTAAATC	-151
N.crassa	AGGACACAAATATCGCAAACCGTGCGGGCGGACGTCCCCGATAGGCTCGG	
MT-L	GAAATGGACGTAATATTCCAGGCAAGAATAATTGTTCGATCAGGAGACAA	-101
N.crassa	ATCGTCGCTAACATGCCATGCCATGCCATGCCATGCCGTGTCGTGGAATG	
MT-L	GTGGGGCATTGGACCGGTTCTTGCAAGCAAGAGCCTATGGCGTGGTGACA	-51
N.crassa	GAAGGTAACAGACAGTCACAGTCACAGTTCCGAACGGTGCGTAGCAGTGT	
MT-L	CGGCGCGTTGCCCATACATCATGCCTCCATCGATGATCCATCCTCACTTG	-1
N.crassa	AGTGTACGTGACAGCAGGAACAAGGTACCGTACGTAGGCATCATTCAGGG	
	+1	
MT-L	**CTATAAAA**AGAGGTGTCCATGGTGCTCAAGCTCAGCCAAGCAAATAAGAC	50
N.crassa	**TATATAAA**GCCACGGGACCCTGACTCCTCCTCAAATTCCTTTACTTGTCA	
MT-L	GACTTGTTTCATTGATTCTTCAAGAGATCGAGCTTCTTTTGCACCACAAG	100
N.crassa	TCAACCGACGAACTAAATCATCACTGCATTGAAACACTCACACAAAGACA	
MT-L	GTCGAGG**ATG**TCTTGCAGCTGCGGATCAAGCTGCGGCTGCGGCTCAAGCT	150
N.crassa	ACATCAGTCTTTCGTCAACCTACACAACCACAACAACTCAATTCCTCTCA	
MT-L	GCAAGTGCGGGTAATATATAATAATATATAAGTGCACCGTGCATGATTAA	200
N.crassa	AAAGAACAACCAGTTCTATCAAA**ATG**GGTGACTGCGGCTGCTCCGGCGCT	

Motif	PsMT	AtMT-1	AtMT-2	MT-L	N. crassa
TCGCCA(N)$_{2-4}$AATTTG	-150 to -135	-189 to -176			
ATCTA	-99	-91			
TGGA	-38	-39			
TCTCTC(N)$_{5}$TATA(N)$_{3}$GGC	-11 to +14	-11 to +14			
CACNCCACAAT	-134		-152		
ACGT	-95		-112		
AGGGT	-79		-90		
GGTCAANATT	-32		-23		
TCTC(N)$_{7}$TATA(N)$_{2}$GGG	-11 to +13		-11 to +13		
ATCATTAA	+22		+24		
AATCG	-152			-154	
AAATGA		-75	-77		
ATGTAT		-32	-29		
TCTC(N)$_{2}$TT(N)$_{3}$TATA(N)$_{3}$GG(N)$_{3}$TTCC		-11 to +20	-11 to +20		
TGATCC		-22		-17	
AAGAAT			-172	-191	
ACTC(N)$_{4}$ATTATT(N)$_{7-12}$TGGAC			-151 to -125	-172 to -142	
GTGACA				-57	-43

Table 2. Location of motifs within the 5' flanking regions of MT-like genes from plants. Distances from the putative TATA box are shown. .

Optimisation of automated DNA sequence analysis.

There are a number of considerations to be taken into account if high-quality DNA sequence data were to be obtained from an automated DNA sequencer. The purity of the DNA template is of the utmost importance. It has been demonstrated, using various DNA preparation methods, that one of the main problems often associated with inferior sequence data (low signal to high background ratio) is attributed to a high level of salt being present in the DNA sample. It is also required that the concentration of the DNA sample has been estimated correctly since both the number of bases obtained in a sequencing run and the accuracy of base-calling are strongly dependent on the amount of template used in the sequencing reaction. These problems may be virtually eliminated when preparing DNA by one of the methods where the nucleic acid is absorbed to a solid support if extra washing of the DNA bound to the resin is carried out, and the DNA is eluted with water at 55 °C - 70 °C instead of the buffers recommended by the manufacturers. This reduces the salt concentration in the DNA sample. However, if the DNA is required for other purposes and long term storage, it is not recommended to store DNA samples in water as the DNA is an acid and will undergo autocatalytic hydrolysis in the absence of a buffering agent. Sequence data obtained from a template contaminated with salt is usually ambiguous with a reduced signal. If there are sufficient amounts of the DNA sample, an ethanol precipitation step with three 70 % ethanol washes has been demonstrated to improve the quality of the sequence data, which is presumed to be due to the removal of contaminating salts. To estimate the concentration of the DNA samples for sequencing, a standard of known concentration was used as a control (usually pGEM 3Zf(+) 200 ng μl^{-1} from Applied Biosystems Ltd. sequencing kits) with agarose gel electrophoresis. When a DNA sample is under estimated (ie too much DNA was added to the reaction) the signal at the start of the sequence is very high, usually going off scale, and diminishes rapidly within the first 250 bases from the start of the sequence and the sequence beyond 250 bases is usually ambiguous with a high background to low signal ratio. This is due to the algorithm assigned to the analysis software which automatically analyses from base one to the end of the sequence. To improve the signal to noise ratio, the start and end points were manually assigned. By analysing the data from base 1 to around base 250 the analysis software was not required to produce unreadable sequence beyond this point, and in turn not over amplify the signal at the beginning of the sequence. When a DNA sample is over estimated (ie insufficient DNA was added to the reaction), and the DNA concentration is only marginally incorrectly estimated, 150-250 bases of readable sequence data can be obtained by manually assigning the start and end points. However, if the DNA concentration is considerably over estimated no readable sequence data will be obtained. Another factor which can affect the quality of the sequence data for automated sequence analysis is the source of the host cells. Highly recommended strains are DH5α and HB101, recommended strains are JM109, XL1Blue and MV1190. However, JM101 is not recommended for use with automated sequence analysis (Taylor *et al*., 1993).

An attempt was made to develop / optimise the methodology for sequencing cosmid clones prepared by alkaline lysis / PEG precipitation with specific primers, as opposed to sub-cloning and sequencing using universal primers. This would have reduced the time required to obtain the sequence data required for gene characterisation. There are still a number of problems associated with this methodology. The sequence data obtained using the alkaline lysis / PEG precipitation method was not acceptable for use as novel sequence data but only as confirmation of homology of the genomic clone to the cDNA sequence. Cosmid DNA was subsequently obtained from a CsCl gradient, however, the sequence data was not of the same high standard as would have been expected if the specific fragment had been cloned into conventional vectors. On the other hand, even if the sequence data had been of an acceptable standard, the time taken to prepare DNA from CsCl gradients, would make this method less efficient than sub-cloning the fragment and using one of the small scale DNA preparations as described in materials and methods.

On comparison of the methods previously described for the purification of PCR products for sequencing, it has been observed that all methods detailed will result in sequence data of a high standard. When purifying products using one of the recommended kits, it is advisable either to increase the volume of buffer used for the washing step and elute in water or to add an ethanol precipitation step at the end to ensure sequence data of a high standard. This will reduce any problems similar to those encountered for plasmid DNA purification and subsequent sequencing.

Future work.

Further analysis of the regulation of different types of MT - like plant genes (eg in *A. thaliana*)

With respect to what is known about MT expression in pea roots, further work is required to characterise the upstream regions of *A. thaliana* genes. There are a number of advantages in using *A. thaliana* to further characterise these regions. *A thaliana* is known as the "botanical *Drosophila* " (Whyte, 1946), due to its relatively small genome size (1.9×10^8 bp) compared to those of pea (5.0×10^9 bp) or maize (3.2×10^9 bp), and its amenability as a genetic system since the mature plant is small and compact and therefore allows large numbers to be grown in a relatively small space, either in sterile media or in soil. The generation time from seed to seed is ca. 6 weeks when grown in continuous illumination, therefore, screening of seedlings is rapid and inexpensive. We potentially have two out of the three probes characterised which will be required to further examine the upstream regions of the *A. thaliana* MT-like genes. The third PCR product amplified from an *A. thaliana* cDNA library (figure 24) requires cloning into the pGem T vector system from Promega and characterisation. Once the sequence has been compared to the cDNA sequence obtained by Raynal *et al* (1993), it can be used to screen an *A. thaliana* genomic library. More restriction analyses of the second hybridising clone (B) from *A. thaliana* type 1 (figure 18) and of clone 4 from *A. thaliana* type 2 will be required to produce smaller fragments for sub-cloning and characterisation.

Cis -acting gene regulatory elements have been found which regulate a variety of environmentally controlled plant genes. There are several ways to locate and identify such upstream regulatory sequences. Once a particular gene has been cloned and characterised, putative regulatory elements can be proposed by comparing sequences of genes which are activated under similar conditions, and identify any consensus sequences. The work reported herein has already identified several 'candidate' sequences within MT-like plant genes (table 2). However, these sequence comparisons must be complemented with functional analysis since homologies identified based on consensus sequences may reflect evolutionary processes of gene duplication rather than functional sequence conservation. A standard experimental procedure involving transgenic plant technology is to construct deletions of the 5' flanking regions of a gene and transfer these modified genes to a host plant where there expression can be easily analysed. The 5' flanking sequences can be systematically deleted either by using existing restriction enzyme sites in the sequence, or by using enzymes such as *Bal* 31 nuclease which hydrolises double-stranded DNA molecules by progressively removing nucleotides from an exposed terminus. In addition putative regulatory elements (for example the identified motifs shown in table 2) may be specifically mutated via PCR-based methods. There are two methods which may be used to investigate the activity of these deleted gene sequences. One method is to use northern blotting for assaying the accumulation of mRNA species or alternatively if antibodies to the specific protein are available the protein product of the genes may be assayed using enzyme-linked immunosorbent assay (ELISA). Once a deletion series has been prepared for a gene under investigation it is cloned back into its host, however, problems may occur as the host will almost certainly possess a copy of the cloned gene. The changes in the expression pattern of the cloned gene will not be distinguished from the normal pattern of expression displayed by the host's copy of the gene.

Alternatively, if an assay for the normal product of a gene is not available, or if it is desired to investigate the precise cell types where a particular gene is being expressed, the coding sequence of a gene can be replaced with that of a "reporter" gene. Such constructs are termed promoter-reporter gene fusions. When cloned into the host organism the expression pattern of the reporter gene will exactly mimic that of the original gene, as the reporter gene will be under the influence of exactly the same control as the original gene. The reporter gene must be chosen with care and the first criterion is that the reporter gene must code for a phenotype not already displayed by the host organism. The phenotype of the reporter gene must be relatively easy to detect after it has been cloned into the host and ideally it should be possible to assay the phenotype quantitatively. Commonly used reporter genes are the bacterial antibiotic resistance genes coding for *cat* and neomycin phosphotransferase (*npt* II), and a third widely used gene is β-glucuronidase (*gus*). If a transgenic plant is expressing either *cat* or *npt* II in a reporter gene fusion, the amounts of the enzyme being produced and therefore the activity of the test promoter can be easily assayed. However, these two systems do not permit the localisation of expression of the fusion genes in particular cell types, since the enzyme assays require the use of homogenised transgenic tissues. A reporter gene which does allow visualisation of expression in particular cell types is *gus* . GUS is a an enzyme from *E. coli* which catalyses the cleavage of a variety of glucuronides, many of which are commercially available as spectrophotometric, fluorometric and histochemical substrates. In the presence of the enzyme GUS the histochemical substrate 'X-gluc' reacts to produce a blue compound which can be easily seen under a microscope and specific cell types expressing a particular promoter-*gus* gene fusion can be identified. This can reveal where promoters of genes of unknown function are expressed.

Deletion analysis is not only used to locate control sequences but importantly can also indicate the function of each sequence. The technique depends on the assumption that deletion of the control sequence will result in a change in the way in which expression of the cloned gene is regulated. For instance, deletion of a sequence that represses or silences expression of a gene should result in that gene being expressed at a higher level, a decrease in expression will indicate that an activator or enhancer has been deleted. Similarly, tissue-specific control sequences can be identified as their deletion will result in the target gene being expressed in tissues other than the correct one. The results of a deletion analysis project have to be interpreted very carefully. Complications may arise if a single deletion removes two closely linked control sequences or, as is fairly common, two distinct control sequences cooperate to produce a single response.

As a first step to characterising the promoters of the MT-like genes from *A. thaliana* , the transcriptional start site for each gene should be determined by methods such as S1 mapping and primer extension. Once this has been done, separate transgenic *A. thaliana* plants with the promoters fused to β-glucuronidase (gus) can be obtained. The developing *A. thaliana* transgenic plants should be monitored for patterns of gus expression in response to different levels of metal ions (zinc, cadmium, iron and copper). A working hypothesis is that i) the E_C type promoter may drive seed specific expression of GUS (and also responsive to ABA), ii) the type 1 promoter may drive Cu/Fe responsive expression in roots of GUS, iii) the type 2 promoter may drive metal responsive expression in shoots of GUS. It will be of interest to establish whether or not if any of the motifs identified in table 2 are in fact involved in conferring these patterns of expression.

Once the promoters have been roughly mapped to identify *cis* -acting elements, DNA / protein interactions can be studied using gel shift assays, to identify these elements to which proteins associate. For example, it could be predicted that binding of specific proteins to upstream regions, at least of type 1's, will be copper dependent and by using gel retardation assays this can be determined. If copper responsive elements are found then it may be possible to clone the protein(s) that interact with them. Specific DNA fragments can be labelled and used as probes to screen cDNA expression libraries. Any expressed proteins that bind specific fragments would be putatively involved in controlling MT gene expression.

Further analysis of the functions of different types of MT - like plant genes (eg in *A. thaliana*)

To examine the metal-binding properties of the putative proteins from *A. thaliana* , similar studies to the ones for pea reported by Tommey *et al* ., (1991); and Robinson *et al* ., (1992) need to be performed. These involved expressing the proteins in *E.coli* as carboxyterminal extensions of glutathione-S-transferase (GST), purification of the proteins from lysates of *E.coli* grown in metal supplemented media (ie zinc, cadmium, iron and copper) and estimation of the pH of half-dissociation of *in vitro* associated metal ions. These results can be compared with those obtained for pea (Tommey *et al* ., 1991; Robinson *et al* ., 1992) and equine MT (Kägi and Vallee, 1960, 1961) to establish whether or not the products of these other gene genes differ in their metal-binding properties. $PsMT_A$ was shown, in vitro, to have a high affinity for copper ions (Tommey *et al* ., 1991). The products of other types of MT-like genes could conceivably show a relatively greater affinity for zinc.

Gene function may be examined using reversed genetics. The cauliflower mosaic virus 35S promoter provides high levels of expression of heterologous genes in a variety of different cell types of many different dicotyledenous and monocotyledenous plant species. The promoter is highly active in both transient expression and stably transformed plant cells. The coding region of a particular plant gene can be associated with the CaMV 35S promoter which will allow over-expression of the gene, or alternatively the coding region can be inserted in the reverse orientation which "should" impart an effect of reduced-expression of the gene. These experiments can be used to increase or decrease the expression of all three types of genes from *A. thaliana* and thereby examine any phenotypic changes. Possible phenotypic changes may include modified seed development with respect to Ec type, modified metal tolerance / accumulation in roots with respect to type 1, modified metal tolerance / accumulation in shoots with respect to type 2.

BIBLIOGRAPHY

Andersen, R.D., Birren, B.W., Taplitz, S.J., Herschman, H.R. (1986) Rat metallothionein-1 structural gene and three pseudogenes, one of which contains 5-regulatory sequences. Mol. Cell. Biol. **6**: 302-314.

Ansorge, W., Sproat, B.S., Stegeman, J., Schwager, C. (1986) A non-radioactive automated method for DNA sequences determination. J. Biochem. Biophys. Meth. **13**: 315-323.

Baker, J., Steel, C., Dure, L. III (1988) Sequence and characterisation of 6 *Lea* proteins and their genes from cotton. Plant Mol. Biol. **11**: 277-291.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for srceening recombinant DNA. Nucl. Acid. Res. **7**: 1513-1523.

Buchman, C., Skroch, P., Welch, J., Fogel, S., Karin, M. (1989) The *CUP2* gene product, regulator of yeast metallothionein expression, is a copper-activated DNA-binding protein. Mol. and Cell. Biol. **9**: 4091-4095.

Butt, T.R., Sternberg, E.J., Gorman, J.A., Clark, P., Hamer, D.H., Rosenberg, M., Crooke, S.T. (1984) Copper metallothionein of yeast, structure of the gene, and regulation of expression. Proc. Natl. Acad. Sci. U.S.A. **81**: 3332-3336.

Butt, T.R. and Ecker, D.J. (1987) Yeast metallothionein and applications in biotechnology. Microbiological Reviews **51**: 351-364.

Carter, A.D., Felber, B.K., Walling, M.J., Jubier, M.F., Schmidt, C.J., Hamer, D.H. (1984) Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Natl. Acad. Sci. U.S.A. **81**: 7392-7396.

Casas-Finit, J.R., Hu, S., Hamer, D.H., Karpel, R.L. (1992) Characterisation of the copper- and silver- thiolate clusters in N-terminal fragments of the yeast ACE1 transcription factor capable of binding to its specific DNA recognition sequence. Biochemistry **31**: 6617-6626.

Chien, A., Edgar, D.B., Trela, J.M. (1976) Deoxyribonucleic acid polymerase from extreme thermophile *Thermus aquaticus* . J. Bacteriol. **127**: 1550-1557

Cimino, G.D., Metchette, K., Isaacs, S.T., Zhu, Y.S. (1990) More false positive problems. Nature **345**: 773-774.

Compere, S.J. and Palmiter, R.D. (1981) DNA methylation controls the inducibility of the mouse metallothionein I gene in lymphoid cells. Cell **25**: 233-240.

Culotta, V.C., Hamer, D.H. (1989) Fine mapping of a mouse metallothionein gene metal response element. Mol. Cell. Biol. **9**: 1376-1380.

Culotta, V.C., Hsu, T., Hu, S., Fürst, P., Hamer, D.H. (1989) Copper and the ACE1 regulatory protein reversibly induce yeast metallothionein gene transcription in a mouse extract. Proc. Natl. Acad. Sci. U.S.A. **86**: 8377-8381.

De Framond, A.J. (1991) A metallothionein-like gene from maize (*Zea mays*): Cloning and characterisation. FEBS Letters **290**: 103-106.

De Miranda, J.R., Thomas, M.A., Thurman, D.A., Tomsett, A.B. (1990) Metallothionein genes from the flowering plant Mimulus guttatus . FEBS Letters **260**: 277-280.

Erlich, H.A., Gelfand, D.H., Saiki, R.K. (1988) Specific DNA amplification. Nature **331**: 461-462.

Evans, I.M., Gatehouse, L.N., Gatehouse, J.A., Robinson, N.J., Croy, R.R.D. (1990) A gene from pea (*Pisum sativum* L.) with homology to metallothionein genes. FEBS Letters **262**: 29-32.

Evans, K.M., Gatehouse, J.A., Lindsay, W.P., Shi, J., Tommey, A.T., Robinson, N.J. (1992) Exprssion of the pea metallothionein-like gene $PsMT_A$ in *Escherichia coli* and *Arabidopsis thaliana* and analysis of trace metal ion accumulation: Implications for $PsMT_A$ function. Plant Mol. Biol. **20**: 1019-1028.

Fordham-Skelton, A.P., Yarwood, A., Croy, R.R.D. (1990) Synthesis of saporin gene probes from partial protein sequence data: Use of inosine-oligonucleotides, genomic DNA and the polymerase chain reaction. Mol. Gen. Genet. **221**: 134-138.

Fowler, B.A., Hindebrand, C.E., Kojima, Y., Webb, M. (1987) Nomenclature of metallothionein. *In* Metallothionein II (eds) Kägi, J.H.R. and Kojima, Y., pp. 19-22. Birkhauser Verlag, Basel.

Furey, W.R., Robins, A.H., Clancy, L.L., Winge, D.R., Wang, B.C., Stout, C.D. (1987) Crystal structure of Cd, Zn. Metallothionein. Exp. Suppl., **52**: 139-148.

Gould, S.J., Subramani, S., Scheffler, I.E. (1989) Use of the DNA polymerase chain reaction for homolgy probing: Isolation of partial cDNA or genomic clones encoding the iron-sulphur protein of succinate dehydrogenase from several species. Proc. Natl. Acad. Sci. U.S.A **86**: 1934-1938.

Grill, E., Winnacker, E.L., Zenk, M.H. (1985) Phytochelatins: The principal heavy-metal complexing peptides of higher plants. Science **230**: 674-676.

Grill, E., Winnacker, E.L., Zenk, M.H. (1987) Phytochelatins, a class of heavy metal-binding peptides from plants, are functionally analogous to metallothioneins. Proc. Natl. Acad. Sci. U.S.A. **84**: 439-443.

Grill, E., Löffler, S., Winnacker, E-L., Zenk, M.H. (1989) Phytochelatins, the heavy -metal binding peptides of plants are synthesised are synthesised from glutathione by a specific g-glutamylcysteine dipeptidyl transpeptidase (phytochelatin synthase). Proc. Natl. Acad. Sci. U.S.A. **86**: 6838-6842.

Guiltinan, M.J., Marcotte, W.R. Jr., Quantrano, R.S. (1990) A plant leucine zipper protein that recognises an Abscisic Acid response element. Science **250**: 267-271.

Gyllensten, U.B. and Erlich, H.A. (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc. Natl. Acad. Sci. U.S.A. **85**: 7652-7656.

Hamer, D.H. (1986) Metallothionein. Annual Review of Biochemistry **55**: 913-951.

Hamer, D.H. and Khoury, G. (1983) *In* Enhancers and Eukaryotic Gene Expression, (eds) Y.Gluzman, T. Shenk, pp1-15. New York: Cold Spring Harbor Lab.

Harlow, P., Watkins, E., Thornton, R.D., Nemer, M. (1989) Structure of an ectodermally expressed sea urchin metallothionein gene and characterisation of its metal responsive region. Mol. Cell Biol. **9**: 5445-5455.

Higham, D.P., Sadler, P.J., Scawen, M.D. (1984) Cadmium-resistant *Pseudomonas putida* synthesises novel cadmium binding proteins. Science **225**: 1043-1046.

Huckle, J.W., Morby, A.P., Turner, J.S., Robinson, N.J. (1993) Isolation of a prokaryotic metallothionein locus and analysis of transcriptional control by trace metal ions. Mol. Microbiol. **7**: 177-187.

Jackson, P.J., Unkefer, C.J., Doolen, J.A., Watt, K., Robinson, N.J. (1987) Poly(g-glutamyl-cysteinyl) glycine: its role in cadmium resistant plant cells. Proc. Natl. Acad. Sci. U.S.A. **84**: 6619-6623.

Jacobs, F.A., Romeyer, F.M., Beauchemin, M., Brousseau, R. (1989). Human metallothionein-II is synthesised as a stable membrane-localised fusion protein in *Escherichia coli* . Gene **83**: 95-103.

Johnson, G.F., Morell, A.G., Stockert, R.J. (1981) Hepatology (Baltimore) I: 243-248.

Kadonaga, J.T., Jones, K.A., Tjian, R. (1986) Trends Biochem. Sci. **11**: 20-23.

Kägi, J.H.R. and Kojima, Y. (1987) Chemistry and biochemistry of metallothionein. *In* Metallothionein II (eds) Kägi, J.H.R. and Kojima, Y., pp. 25-61. Birkhauser Verlag, Basel.

Kägi, J.H.R. and Schäffer, A. (1988) Biochemistry of metallothionein. Biochemistry **27**: 8509-8515.

Kägi, J.H.R. and Vallee, B.L. (1960) Metallothionein: A cadmium- and zinc-containing protein from the equine kidney cortex. J. Biol. Chem. **235**: 3460-3465.

Kägi, J.H.R. and Vallee, B.L. (1961) Metallothionein: A cadmium and zinc-containing protein from the equine kidney cortex. J. Biol. Chem. **236**: 2435-2442.

Kambara, M., Nishikawa, T., Katayama, Y., Yamaguchi, T. (1988) Optimisation of parameters in a DNA sequencer using fluorescence detection. Bio/Technology Vol. **6** Iss. 7: 816-821.

Karin, M. and Richards, R.I. (1982) Human metallothionein genes: Molecular cloning and sequence analysis of the mRNA. Nucl. Acid. Res. **10**: 3165-3173.

Karin, M., Najarian, R, Haslinger, A., Valenzuela, P., Welch, J., Fogel, S. (1984) Primary structure and transcription of an amplified genetic locus: The CUP1 locus of yeast. Proc. Natl. Acad. Sci. U.S.A. **81**: 337-341.

Karin, M., Haslinger, A., Holtgreve, H., Richard, R.I., Krailter, P., Westphal, H.M., Beato, M. (1984) Characterisation of DNA sequences through cadmium and glucocorticoid hormones induce human MT-II_A gene. Nature **308**: 513-519.

Karin, M., Haslinger, A., Heguy, A., Dietlin, T., Cooke, T. (1987) Metal-responsive elements act as positive modulators of human metallothionein II_A enhancer activity. Mol. Cell Biol. **7**: 606-613.

Karlovsky, P. (1990) Misuse of PCR. TIBS. **15**: 419.

Kawashima, I., Inokuchi, Y., Chino, M., Kimura, M., Shimizu, N. (1991) Isolation of a gene for a metallothionein-like protein from soybean. Plant Cell Physiol. **32**: 913-916.

Kawashima, I., Kennedy, T.D., Chino, M., Lane, B.G. (1992) Wheat E_c metallothionein genes: Like mammalian Zn^{2+} metallothionein genes, wheat Zn^{2+} metallothionein genes are conspicuously expressed during embryogenesis. Eur. J. Biochem. **209**: 971-976.

Kay, J., Cryer, A., Darke, B.M., Kille, P., Lees, W.E., Norey, C.G., Stark, J.M. (1990) Int. J. Biochem. **23**: 1-5.

Keohavong, P., Kat, A.G., Cariello, N.F., Thilly, W.G. (1988) DNA amplification using T4-DNA polymerase. DNA (N Y) Vol. **7**. Number 1: 63-70.

Kille, P., Stephens, P., Cryer, A., Kay, J. (1990) The expression of a synthetic rainbow trout metallothionein gene in *E.coli* . Biochem. Biophys. Acta. **1048**: 178-186.

Kille, P., Winge, D.R., Harwood, J.L., Kay, J. (1991) A plant metallothionein produced in *E.coli* . FEBS Lett. **295**: 171-175.

Kitchen, P.A., Szotyori, Z., Fromholc, C., Almond, N. (1990) Avoidance of false positives. Nature **344**: 201.

Klemsdal, S.S., Hughes, W., Lönneborg, A., Aalen, R.B., Olsen, O-D. (1992) Primary structure of a novel barley gene differentially expressed in immature aleurone layers. Mol. Gen. Genet. **228**: 9-16.

Kwok, S. and Higuchi, R. (1989) Avoiding false positives with PCR. Nature **339**: 237-238.

Labbé, S., Prévost, J., Remondelli, P., Leone, A., Séquin, C. (1991) A nuclear factor binds to the metal regulatory elements of the mouse gene encoding metallothionein-1. Nucl. Acid. Res. **19**: 4225-4231.

Labbé, S., Larouche, L., Mailhot, D., Séquin, C (1993) Purification of mouse MEP-1, a nuclear protein which binds to the metal regulatory elements of genes encoding metallothionein. Nucl. Acid. Res. **21**: 1549-1554.

Lane, B.G., Kajioka, R., Kennedy, T.D. (1987) The wheat-germ E_C protein is a zinc-containing metallothionein. Biochem. Cell Biol. **65**: 1001-1005.

Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M., Caskey, C.T. (1988) Generation of cDNA probes directed by amino acid sequence: Cloning of urate oxidase. Science **239**: 1288-1291.

Lerch, K. and Beltramini, M. (1983) Neurospora copper metallothionein: molecular structure and biological significance. Chem. Scripta **21**: 109-115.

Mack, D.H. and Sninsky, J.J. (1988) A sensitive method for the identification of uncharacterised viruses related to known virus groups: Hepadnavirus model system. Proc. Natl. Acad. Sci. U.S.A. **85**: 6977-6981.

MacLean, F.I., Lucis, O.J., Shakh, Z.A., Jansz, E.R. (1972) The uptake and sub-cellular distribution of Cd and Zn in microorganisms. Fed. Proc. **31**: 699.

Margoshes, M. and Vallee, B.L. (1957) A cadmium protein from equine kidney cortex. J. An. Chem. Soc. **79**: 4813-4814.

Maroni, G., Otto, E., Lastowski-Perry, D. (1986) Molecular and cytogenic characterisation of a metallothionein gene of *Drosophila* . Genetics **112**: 493-504.

Masters, B.A., Kelly, E.J., Quaife, C.J., Brinster, R.L. (1994) Targeted disruption of metallothionein I and II genes increases sensitivity to cadmium. Proc. Natl. Acad. Sci. U.S.A. **91**: 584-588.

Maxim, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**: 560-564.

Mayo, K.E. and Palmiter, R.D. (1981) Glucocorticoid regulation of metallothionein-I mRNA synthesis in cultured mouse cells. J. Biol. Chem. **256**: 2621-2624.

Mayo, K.E., Warren, R., Palmiter, R.D. (1982) The mouse metallothionein-I gene is transcriptionally regulated by cadmium following transfection in Human or Mouse cells. Cell **29**: 99-108.

Morby, A.P., Turner, J.S., Huckle, J.W., Robinson, N.J. (1993) SmtB is a metal regulated repressor of the cyanobacterial metallothionein gene *smtA* : identification of a Zn inhibited DNA-protein complex. Nucl. Acid. Res. **21**: 921-925.

Mueller, P.R., Salser, S.J., Wold, B. (1988) Constitutive and metal-inducible protein: DNA interactions of the mouse metallothionein I promoter examined by *in vivo* and *in vitro* footprinting. Genes Dev. **2**: 412-427.

Mullis, K.B. and Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**: 335-350.

Mullis, K.B., Faloona, F., Scarf, S., Saiki, R.K., Horn, G.T., Erlich, H.A. (1986) Specific enzymatic amplification of DNA *in vitro* : The polymerase chain reaction. Cold Spring Harbor Symp. Quant. Biol. **51**: 263-274.

Mundy, J., Yamaguchi-Shinozaki, K., Chua N-H. (1990) Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. Proc. Natl. Acad. Sci. U.S.A. **87**: 1406-1410.

Münger, K., Germann, U.A., Lerch, K. (1987) The Neurospora crassa metallothionein gene. J.Biol. Chem. **262**: 7363-7367.

Münger, K. and Lerch, K. (1985) Copper metallothionein from the fungus *Agaricus bisporus* : chemical and spectroscopic properties. Biochem. **24**: 6751-6756.

Murasugi, A., Wada, C., Hayashi, Y. (1981) Purification and unique properties in UV and CD spectra of Cd-binding peptide 1 from *Schizosaccharomyces pombe* . Biochem. Biophy. Res. Comm. **103**: 1021-1024.

Murooka, Y. and Nagaoka, T. (1987) Expression of cloned monkey metallothionein in *Escherichia coli* . Appl. Environ. Microbiol. **53**: 204-207.

Nartey, N.O., Banerjee, D., Cherian, M.G. (1987) Immunohistochemical localisation of metallothionein in cell nucleus and cytoplasm of fetal human liver and kidney and its changes during development. Pathology **19**: 233-238.

Nielson, K.B., Winge, D.R. (1983) Order of metal-binding in metallothionein. J. Biol. Chem. **258**: 13063-13069.

Nemer, M., Wilkinson, D.G., Travaglini, E.C., Sternberg, E.J., Butt, T.R. (1985) Sea Urchin metallothionein sequence: key to an evolutionary diversity. Proc. Natl. Acad. Sci. U.S.A. **82**: 4991-4994.

Ochman, H., Gerber, A.S., Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. Genetics **120**: 621-623.

Okumura, N., Nishizawa, N-K., Umehara, Y., Mori, S. (1991) An iron deficiency-specific cDNA from barley roots having two homologous cysteine-rich MT domains. Plant Mol. Biol. **17**: 531-533.

Olafson, R.W., Abel, K., Sim, R.D. (1979) Prokaryotic metallothionein: preliminary characterisation of a blue green alga heavy metal binding protein. Biochem. and Biophys. Res. Comm. **89**: 36-43.

Olafson, R.W., Loya, S., Sim, R.D. (1980) Physical parameters of prokaryotic metallothionein induction. Biochem. and Biophys. Res. Comm. **95**:1495-1503.

Olafson, R.W. (1984) Prokaryotic metallothionein. Int. J. of Peptide and Protein Res. **24**: 303-308.

Olafson, R.W., McCubbin, W.D., Kay, C.M. (1988) Primary and secondary-structural analysis of a unique prokaryotic metallothionein from a *Synechococcus* sp. cyanobacterium. Biochem. Journ. **251**: 691-699.

Olafson, R.W. (1991) Purification of prokaryotic metallothioneins. Meth. Enzymol. **205**: 283-286.

Olsen, O-A., Jakobsen, K.S., Schmelzer, E. (1990) Development of barley aleurone cells: Temporal and spatial patterns of accumulation of cell-specific mRNAs. Planta **181**: 462-466.

Otto, E., Allen, J.M., Young, J.E., Palmiter, R.D., Maroni, G. (1987) A DNA segment controlling metal-regulated expression of the *Drosophila melanogaster* metallothionein gene *Mtn* . Mol. Cell. Biol. **7**: 1710-1715.

Palmiter, R.D. (1987) Molecular biology of metallothionein gene expression. *In* Metallothionein II (eds) Kägi, J.H.R. and Kojima, Y., pp. 63-80. Birkhauser Verlag, Basel.

Palmiter, R.D., Findley, S.D., Whitmore, T.E., Durnam, D.M. (1992) MT-III, a brain-specific member of the metallothionein gene family. Proc. Natl. Acad. Sci. U.S.A. **89**: 6333-6337.

Premakuman, R., Winge, D.R., Wiley, R.D. Rajagophan, K.V. (1975) Copper-chelatin: Isolation from various eukaryotic sources. Arch. Biochem. Biophys. **170**: 278-288.

Prober, J.M., Trainer, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jenson, M.A., Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science **238**: 336.

Radtke, F., Heuchel, R., Georgiev, O., Hergersberg, M., Gariglio, M., Dembic, Z., Schaffner,W. (1993) Cloned transcription factor MTF-1 activates the mouse metallothionein-1 promoter. The EMBO Journal. **12**: 1355-1362.

Rauser, W.E. (1990) Phytochelatins. Ann. Rev. Biochem. **59**: 61-68.

Raynal, M. Grellet, F., Laudie, M., Meyer, Y., Cooke, R., Delseny, M. (1993) OWL: accession number Z27049.

Reece, R.N., Mehra, R.K., Tarbet, E.B., Winge, D.R. (1988) Studies on the gamma-glutamyl Cu-binding peptide from *Schizosaccharomyces pombe* . J. Biol. Chem. **263**: 4186-4192.

Richards, R.I., Hugey, A., Karin, H. (1984) Structural and functional analysis of the human MT-1_A gene: differentail induction by metal ions and glucocorticoids. Cell **37**: 263-272.

Robinson, N.J., Jackson, P.J. (1986) 'Metallothionein like' metal complexes in angiosperms; their srtucture and function. Plant Physiol. **67**: 499-506.

Robinson, N.J., Ratcliff, R.L., Anderson, P.J., Delhaize, E., Berger, J.M., Jackson, P.J. (1988) Biosynthesis of poly(gamma-Glutamylcysteinyl)glycines in cadmium-resistant *Datura innoxia* cells. Plant Sci. **56**: 197-204.

Robinson, N.J. (1990) Metal binding polypeptides in plants. In Heavy metal tolerance in plants: Evolutionary aspects (ed.) Shaw, J.A., pp. 195-213. CRC Press, Florida.

Robinson, N.J., Gupta, A., Fordham-Skelton, A.P., Croy, R.R.D., Whitton, B.A., Huckle, J.W. (1990) Prokaryotic metallothionein gene characteristion and expression: Chromosome crawling by ligation-mediated PCR. Proc. R. Soc. Lond. B. **242**: 241-247.

Robinson, N.J., Evans, I.M., Mulcrone, J., Bryden, J., Tommey, A.M. (1992) Genes with similarity to metallothionein genes and copper, zinc ligands in *Pisum sativum* L. Plant and Soil. **146**: 291-298.

Robinson, N.J., Tommey, A.M., Kuske, C., Jackson, P.J. (1993) Review article: Plant metallothioneins. Biochem. J. **295**: 1-10.

Romeyer, F.M., Jacobs, F.A., Masson, L., Hanna, Z., Brousseau, R. (1988) Bioaccumulation of heavy metals in *Escherichia coli* expressing an inducible synthetic human metallothionein gene. J Biotechnol. **8**: 207-220.

Romeyer, F.M., Jacobs, F.A., Brousseau, R. (1990) Expression of a *Neurospora crassa* metallothionein and its variants in *Escherichia coli* . Appl. Environ. Microbiol. **56**: 2748-2754.

Saiki, R.K., Scarf, S., Falona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N. (1988) Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of of sickle cell anaemia. Science **230**: 1350-1354.

Sambrook, J., Fritsch, E.F., Maniatus, T. (1989) Recovery and purification of DNA fractionated on agarose gels. *In* Molecular Cloning: A Laboratory Manual, 2nd Edition. Cold Spring Harbor Laboratory, p6.22.

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**: 5463-5467.

Sarkar, G. and Sommer, S. (1990) Shedding light on PCR contamination. Nature **343**: 27.

Scarf, S., Horn, G.T., Erlich, H.A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science **233**: 1076-1078.

Scholer, H., Haslinger, A., Heguy, A., Holtgreve, H., Karin, M. (1986) *In vivo* competition between a metallothionein regulatory element and the SV40 enhancer. Science **232**: 76-80.

Searle, P.F., Stuart, G.W., Palmiter, R.D. (1985) Metal regulatory elements of the mouse metallothionein-I gene. *In* Metallothionein II (eds) Kägi, J.H.R. and Kojima, Y., pp. 407-414. Birkhauser Verlag, Basel.

Searle, P.F. (1990) Zinc dependent binding of a liver nuclear factor to metal response element MRE-a of the mouse metallothionein-1 gene and variant sequences. Nucl. Acid. Res. **18**: 4683-4690.

Séguin, C., Felber, B.K., Carter, A.D., Hamer, D.H. (1984) Competition for cellular factors that activate metallothionein gene transcription. Nature **312**: 781-785.

Séguin, C., Prévost, J. (1988) Detection of a nuclear protein that interacts with a metal regulatory element of the mouse metallothionein 1 gene. Nucl. Acid. Res. **16**: 10547-10560.

Séguin, C. (1991) A nuclear factor requires Zn2+ to bind a regulatory *MRE* element of the mouse gene encoding metallthionein-1. Gene **97**: 295-300.

Shi, J., Lindsay, W.P., Huckle, J.W., Morby, A.P., Robinson, N.J. (1992) Cyanobacterial metallothionein gene expressed in *Escherichia coli* -metal binding properties of the expressed protein. FEBs Lett. **303**: 159-163.

Shimizu, T., Hiyama, T., Ikeuchi, M., Inoue, Y. (1992) Nucleotide sequence of a metallothionein gene of the thermophilic cyanobacterium *Synechococcus vulcanus* . Plant Mol. Biol. **20**: 565-567.

Shyamala, V. and Ames, G.F-L. (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. Gene **84**: 1-8.

Silar, P. and Wegnez, M. (1990) Expression of the *Drosophila melanogaster* metallothionein genes in yeast. FEBS Lett. **269**: 273-276.

Silver, J. and Keerikatte, V. (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. **63**: 1924-1928.

Smith, L.M., Fung, S., Hunkapiller, M.W., Hunkapiller, T.J., Hood, L.E. (1985) The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus; synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic Acids Res. **13**: 2399-2412.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. Nature **321**: 674-679.

Snowdon, K.C., Gardner, R.C. (1993) Five genes induced by aluminium in wheat (Triticum aestivum L.) roots. Plant Physiol. **103**: 855-861.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by electrophoresis. J. Mol. Biol. **98**: 503-517.

Steffens, J.C. (1990) The heavy metal-binding peptides of plants. Plant Mol. Biol. **41**: 553-575.

Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L., Palmiter, R.D. (1984) A 12-base pair DNA motif that is repeated several time in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci. U.S.A. **81**: 7318-7322.

Stuart, G.W., Searle, P.F., Palmiter, R.D. (1985) Identification of multiple metal regulatory elements in mouse metallothionein-I promoter by assaying synthetic sequences. Nature **317**: 828-831.

Takahashi, K. (1991) OWL: accession number X62818, EMBL, Heidelberg, Germany.

Takatera, K. and Watanabe, T. (1992) Application of high performance liquid chromatography/inductively coupled plasma mass spectrometry to the speciation of cadmium-binding metallothionein-like protein in a cyanobacteria. Anal. Sc. **8**: 469-474.

Taylor, R.G., Walker, D.C., McInnes, R.R. (1993) *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. Nucl. Acid. Res. Vol. **21**: 1677-1678.

Thiele, D.J. and Hamer, D. (1986) Tandemly duplicated upstream control sequences mediate copper-induced transcription of the *Saccharomyces cerevisiae* copper-metallothionein gene. Mol. and Cell. Biol. **6**: 1158-1163.

Thiele, D.J. (1992) Metal-regulated transcription in eukaryotes. Nucl. Acid. Res. **20**: 1183-1191.

Tirglia, T., Peterson, M.G., Kemp, D.J. (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundries of known sequences. Nucl. Acid. Res. **16**: 8186.

Tommey, A.T., Shi, J., Lindsay, W.P., Urwin, P.E., Robinson, N.J. (1991) Expression of the pea gene $PsMT_A$ in E.coli Metal-binding properties of the expressed protein. FEBS **292**: 48-52.

Tsujikawa, K., Imai, T., Katutani, M., Kayamori, Y., Mimura, T., Otaki, N., Kimura, M., Fukuyama, R., Shimizu, N. (1991) Localisation of metallothionein in nuclei of growing primary cultured adult rat hepatocytes. FEBS Lett. **283**: 239-242.

Turner, J.S. (1993) Functional analysis of the prokaryotic metallothionein locus, *smt*. Ph. D. Thesis.

Turner, J.S., Morby, A.P., Whitton, B.W., Gupta, A., Robinson, N.J. (1993) Construction and characterisation of Zn^{2+} / Cd^{2+} hypersensitive cyanobacterial mutants lacking a functional metallothionein locus. J. Biol. Chem. **268**: 4494-4498.

Turner, J.S., Robinson, N.J. (Review in press) Cyanobacterial metallothioneins: Biochemistry and molecular genetics.

Vasak, M. and Armitage, I. (1986) Environ. Health. Perspect. **65**: 215-216.

Weig, A. and Komor, E. (1992) Sequence analysis of a cDNA clone for metallothionein-I-like protein from *Ricinus communis* L. OWL: accession number L02306, EMBL, Heidelberg, Germany.

Welch, J. Fogel, S., Buchman, C., Karin, M. (1989) The *CUP2* gene product regulates expression of the *CUP1* gene, coding for yeast metallothionein. The EMBO J. **8**: 255-260.

Westin, W., Schaffner, W. (1988) A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-1 gene. EMBO Journal **7**: 3763-3770.

Whyte, R.O. (1946) Crop Production and Environment. (London: Faber and Faber).

Zafarullah, M., Bonham, K., Gedamu, L. (1988) Structure of the rainbow trout metallothionein B gene and characterisation of its metal-responsive region. Mol. Cell. Biol. **8**: 4460-4476.

Zhen, L. and Swank, R.T. (1993) A simple and high yield method for recovering DNA from agarose gels. BioFeedBack, BioTechniques Vol **14**: No. 6. 894-896.

Zhou, J. and Goldsbrough, P.B. (1993) OWL: accession number L15389, GENBANK,

Plant and Soil **146**: 291–298, 1992.

PLSO SV91

Genes with similarity to metallothionein genes and copper, zinc ligands in *Pisum sativum* L.

NIGEL J. ROBINSON, I. MARTA EVANS, JANET MULCRONE, JULIA BRYDEN and ANDREW M. TOMMEY
Department of Biological Sciences, University of Durham, Durham DH1 3LE, UK

Key words: copper metabolism, metallothionein, *Pisum sativum* L., *PsMT* genes, zinc metabolism

Abstract

The *PsMT* gene family of pea (*Pisum sativum* L.) encodes predicted proteins with sequence similarity to metallothioneins. However, PsMT proteins have not yet been characterised in planta and their functions remain obscure. *PsMT* transcripts were identified in the cortex tissue of pea roots using tissue squash-blotting techniques. Transcripts were not detected on northern blots of RNA isolated from the embryonic radicle, but *PsMT* transcript abundance in roots increased with age of germinating seedlings. The $PsMT_A$ gene was expressed in *E. coli* as a carboxyterminal extension of glutathione-S-transferase (GST). Fusion protein purified from crude cell lysates (500 mL cultures) bound an estimated amount of 5.99, 6.27 and 7.07 moles of Zn, Cu and Cd respectively per mole protein, compared to equivalent estimates of 0.37, 0.63 and 0.26 moles for GST alone. Similar estimates for Fe-binding were 0.28 moles for GST-$PsMT_A$ fusion protein and 0.1 moles for GST alone.

In summary, these data: 1, show that *PsMT* transcripts are abundant in roots of pea plants that have not been exposed to supra-optimal concentrations of trace metals and hence appear to be constitutively expressed and 2, indicate that $PsMT_A$ protein can bind certain trace metal ions. We have also identified and partially purified a Zn ligand (Zn-A) and two Cu ligands (Cu-A, Cu-B) from pea roots which have not been exposed to supra-optimal conditions of trace metal ions and are therefore defined as 'constitutive'. Whether or not these ligands include the products of *PsMT* genes remains to be established.

Introduction

Metallothioneins, MTs have been isolated from a wide range of vertebrates, invertebrates and micro-organisms. Proposed functions for these cysteine-rich proteins include roles in essential trace metal ion homoeostasis and the detoxification of certain non-essential metals (cited in Kägi and Schäffer, 1988). In animals, MT is most abundant in parenchymatous tissues (liver, kidney, pancreas and intestine) and its abundance varies with age, stage of development and diet. A Zn-binding protein has been isolated from wheat germ (Lane et al., 1987) and designated class II MT (Kägi and Schäffer, 1988) but equivalent proteins have not yet been purified from vegetative plant tissues. Therefore, any proposed role for MTs (class I or II) in essential trace metal nutrition or the detoxification of non-essential metals in vegetative plant tissue is hypothetical.

We have isolated cDNAs (pPR179, pPR705) and a gene, $PsMT_A$, from pea which encode proteins with sequence similarity to MTs (Evans et al., 1990). In addition, we have recently identified two other members of the *PsMT* gene family via PCR-mediated cloning. These genes have been designated $PsMT_B$ and $PsMT_C$. Related cDNAs have been independently isolated from cDNA libraries prepared from mRNA isolated from *Mimulus guttatus* (de Miranda et al., 1990). Homologous gene sequences have also been identified in several other higher plant species. The sequence of the Zn-binding protein from wheat germ (class II MT) does not closely align with the predicted products of these novel plant genes or with known class I MTs. It remains to be established whether or not the products of these novel genes bind metals in planta and their functions are therefore uncertain.

MT genes typically possess metal-regulatory elements. MT transcript abundance increases after exposure to elevated concentrations of certain metal ions in a range of eukaryotes (cited in Palmiter, 1987) and this has also been observed in a prokaryote (Robinson et al., 1990). However, in roots of the monkey flower plant, *M. guttatus*, no increase in the abundance of transcripts encoded by 'MT-like' genes was detected following exposure to elevated concentrations of Cu, Cd or Zn (de Miranda et al., 1990). In addition, a related cDNA (*ids-1*) has recently been isolated from barley, following specific hybridisation to a probe prepared from mRNA isolated from Fe-deficient roots (Okamura et al., 1991). This implies that the *ids-1* sequence is preferentially expressed in barley roots under Fe-limiting conditions. Upon exposure to elevated concentrations of trace metals, higher plants synthesize and accumulate increased amounts of the metal-binding polypeptide poly (γ-glutamylcysteinyl)glycine (Jackson et al., 1987), most commonly known as phytochelatin (Grill et al., 1985) but also referred to as cadystin (Kondo et al., 1985) and class III MT (Kägi and Schäffer, 1988). It has been proposed that phytochelatin and MT (class I) perform analogous functions in plants and animals exposed to elevated concentrations of certain trace metal ions (Grill et al., 1987). It is possible to rationalise that any MT gene present in organisms containing phytochelatins may not show a greatly elevated level of expression in response to supra-optimal concentrations of those metal ions that are efficiently detoxified by phytochelatins.

Animal MT genes additionally show programmed expression during embryogenesis and in different stages of fetal and perinatal development (cited in Kägi and Schäffer, 1988). For example, the hepatic concentration of MT, primarily Zn-thionein, is 20-fold greater in neonatal than in adult rats. The role of MT in these processes is uncertain, although it has been noted that changes in MT concentration coincide with changes in metabolic processes which require trace metals such as Zn. Moreover, there are reports of differential activation of MT isoforms during these processes. Accumulation of MT in nuclei of growing primary cultured-adult-rat hepatocytes has also been observed in early S phase, but not in G_0 or G_1 when the protein is found in the cytoplasm (Tsujikawa et al., 1991). In wheat, class II MT is present in dry embryos but its abundance declines immediately following germination (Hanley-Bowdoin and Lane, 1983). Lane et al. (1987) noted an analogy between wheat and animal MT production during embryogenesis and observed that deposition of a Zn MT during plant embryogenesis may be associated with a shift between proliferative and differentiating stages of development. A detailed description of the pattern of expression of $PsMT_A$, and its homologues, is clearly required.

We report here a preliminary examination of the temporal and spatial pattern of expression of *PsMT* genes in developing pea and provide evidence that the $PsMT_A$ protein is capable of binding certain metal ions. Three metal-complexes have also been identified in pea roots grown in the absence of supra-optimal concentrations of trace metal ions and these may include candidates for the products of *PsMT* genes.

Materials and methods

Plant material

Seeds of *P. sativum* L. (cv Feltham First) were surface-sterilized with 1% v/v Chloros, germinated in the dark for 4 days and grown as previously described by Evans et al. (1990). After a further 9 days, leaves were harvested directly into liquid nitrogen and stored at −80°C. Etiolated leaf material was obtained from plants grown for 4 days in the presence of light and then transferred into the dark for the final 5 days. Roots were harvested from pea plants grown in hydroponic cultures, as described previously (Evans et al. 1988). "Mid-development" cotyledons were obtained from plants grown in hydroponic culture and pods were harvested 15 days after flowering (the period from flowering to seed maturity under these conditions was ca. 22 days). Cotyledons, ascepticallv removed from the testa and embryonic axes, were frozen in liquid nitrogen and stored at −80°C. Radicles were dissected from seeds obtained from pods harvested 13 days after flowering.

Isolation of mRNA, northern and squash blotting

Total RNA and poly(A)$^+$RNA were isolated from different organs (radicle, 4-day-old root, 14-day-old root, developing cotyledon, etiolated leaf, green leaf) as previously described (Evans et al., 1988). Total RNA (5 or 10 μg/lane) and poly(A)$^-$RNA (2 μg/lane) was glyoxalated and electrophoresed through 1.5% agarose gels (McMasters and Carmichael, 1977).

Northern blot analyses were carried out as described previously (Evans et al., 1990). Northern blots were incubated with ^{32}P-labelled $PsMT_A$ cDNA excised from plasmid pPR179 and labelled by random oligonucleotide-priming (Feinberg and Vogelstein, 1983). Filter hybridisation and post-hybridisation washes were carried out under standard conditions.

Squash blotting involved gently pressing freshly excised root material onto a nitrocellulose filter that was previously equilibrated for 15 min with buffer containing 5 × SSC (1 × SSC = 0.15 *M* sodium chloride, 0.015 sodium citrate, pH 7.0), 0.1% w/v SDS (pH 7.5). The nitrocellulose was baked at 80°C under vacuum for 2 h and then incubated with buffer (10 m*M* Tris HCl, pH 7.8; 50 mM EDTA; 0.5% w/v SDS) containing 0.5 mg mL^{-1} proteinase K for 2 h at 37°C prior to hybridization with the ^{32}P-labelled cDNA probe (described above) in a solution containing 50% v/v formamide, 5 × SSC, 5 × Denhardts reagent, 100 μg mL^{-1} denatured herring sperm DNA, 0.1% SDS at 42°C for 48 h. Final post-hybridisation washes were in 1 × SSC, 0.1% SDS at 65°C.

Analysis of metal-binding to $PsMT_A$ protein expressed in E. coli

The $PsMT_A$ protein coding region was cloned into the vector pGEX3X to facilitate expression of the $PsMT_A$ protein in *E. coli* as a carboxyterminal extension of glutathione-S-transferase (GST) as described by Tommey et al. (1991). The fusion protein (M_r 34, 500) was purified from cell lysates using glutathione affinity chromatography (Smith and Johnson, 1988). GST (M_r 26, 500) was similarly purified from lysates of cells transformed with the pGEX3X vector alone. Samples (total volume 2.5 mL) containing either fusion protein or GST were purified from crude lysates of 500 mL cultures grown for 4 h in media supplemented with 500 μM of either Cd, Zn, Cu or Fe. Cd and Zn were added as metal chloride, Cu as metal sulphate. Fe was supplied as Fe citrate, which is available for *fec*-mediated uptake by *E. coli* (Pressler et al., 1988). These samples were passed through Sephadex G-25 (PD-10, Pharmacia) and protein content of the void fraction was estimated using a Coomassie blue based reagent (Bio-Rad) and bovine serum albumin as standard. Metal content was determined by atomic absorption spectrophotometry.

Identification of phytochelatins by reversed phase HPLC

Pea seedlings were grown hydroponically with a final 60 h of growth in either the presence or absence of 50 μM Cd. Extracts were prepared for reversed phase HPLC analysis of phytochelatins by methods previously used for extracts from plant cell suspension cultures (Delhaize et al., 1989) with the following modifications. Harvested roots (between 0.5 and 1.0 g) were washed three times in an excess of distilled water and once with an equal volume of extraction buffer before an equal volume of 1 *N* HCl was added. Tissue was homogenised (Polytron) and cell debris removed by centrifugation (10 000 *g* for 10 min) and also by passage through a 0.2 μm porosity polycarbonate membrane. Samples were then passed through Centricon filtration units (Amicon Corporation, >30,000 Da exclusion), fractionated by reversed phase HPLC and analysed for thiols as described previously (Delhaize et al., 1989).

Purification of constitutive Cu and Zn ligands from pea roots

Plants were grown hydroponically in a 0.05% w/v Phostrogen solution (initial concentrations of Cu and Zn were 0.6 and 0.2 μM respectively, which are assumed to be further depleted during growth) with no additional trace metal ions. Harvested roots (between 150 and 200 g) were washed three times in a large volume of distilled water and then mixed with an equal volume of buffer (0.1 M ammonium acetate, pH 5.5, 1% v/v β-mercaptoethanol) prior to homogenisation (Polytron). The homogenate was strained through two layers of muslin and the resulting solution centrifuged (2 500 g for 15 min). The supernantant was mixed with 4 volumes of acetone, incubated at −80°C for 3 h followed by centrifugation (2 500 g for 5 min). Pellets were resuspended in 50 mL of buffer (10 mM Tris-HCl, pH 7.2, 1% v/v β-mercaptoethanol) heated (60°C for 10 min; insensitivity to such thermal treatment has previously been used to purify MTs from other organisms) and heat denatured debris removed by further centrifugation (2 500 g for 5 min). Samples were passed through Sephadex G-25 (PD-10, equilibrated with 10 mM Tris-HCl pH 7.2, 1% v/v β-mercaptoethanol) and the void fractions loaded onto a column (10 mL) of DEAE Sephadex equilibrated in the same buffer. The matrix was washed with 200 mL of the same buffer followed sequentially by 200 mL of equivalent buffers containing 100 mM and 300 mM Tris-HCl. Fractions (10 mL) were analysed for Cu and Zn by atomic absorption spectrophotometry. Acetone precipitates of pooled fractions containing the major Cu and Zn peaks, prepared as described above, were resuspended in 5 mL of buffer (10 mM Tris-HCl, pH 7.2, 1% v/v β-mercaptoethanol) and fractionated on Sephadex G-50 equilibrated with the same buffer. Every third fraction (2.5 mL) was analysed for Cu, Zn and protein. The void and total volumes of the column were estimated by separation of samples containing blue dextran and free metal ions respectively.

Results and discussion

Figure 1 shows the predicted products of $PsMT_A$ and related plant genes. There is significant sequence similarity within N- and C-terminal cysteine-rich domains which have previously been aligned with know class I MTs (for examples of alignments see Evans et al., 1990; de Miranda et al., 1990). These MT-like domains are separated by a more divergent intervening region.

$PsMT_A$ probe hybridized to transcripts (ca. 640 bases) which were relatively abundant in

```
Tobacco NtMT[a]      --------------------------------------------------------------AAEGGN-GCKCGSNCTCDPCNC-
Alfalfa MsMTA[a]     ---------------------------------------TVIL--GVGPAKIHF-EGAEMGVAAEDG--GCKCGDSCTCDPCNCK
Pea PsMTA[b]         MS---G-CGCGSSCNCGDSCK-CNKRSSGLSYSEMETTETVIL--GVGPAKIQF-EGAEMSAASEDG--GCKCGDNCTCDPCNCK
Pea PsMTB[a]         MS---G-CGCGSSCNCGDSCK-CNKRSSGLSYSEMETTETVIL--GVGPAKIQF-NGAEMSVAAEDG--GCKCGDSCTCDPCNCK
Pea PsMTC[a]         MS---G-CGCGSSCNCGDSCK-CSKRSSGLSYSEMETTETVIL--GVGPAKIQF-NGAEMSVAAEDG--GCKCGDSCTCDPCNCK
M. guttatus[c]       MS--SG-CSCGSGCKCGDNCS-C-SMYPDME-TN--TTVTMIE--GVAPLKM-YSEGSEKSFGAEGGN-GCKCGSNCKCDPCNC-
Soybean[d]           MSCCGGNCGCGSSCKCGNGCGGC-KMYPDLSYTESTTTETLVM--GVAPVKAQF-EGAEMGVPAEND--GCKCGPNCSCNPCTCK
Maize MT-1[e]        MS-----CSCGSSCGCGSSC-KCGKKYPDLEETS--TAAQPTVVLGVAPEKKAAPEFVEAAAESGEAAHGCSCGSGCKCDPCNC-
Barley ids-1[f]      MS-----CSCGSSCGCGSNC-NCGKMYPDLEEKSGATMQVTVIVLGVGSAK----VQFEEAAEFGEAAHGCSCGANCKCNPCNC-

CONCENSUS            MS-----C-CGS-C-CG--C--C------------T--------GV---K-------E---------GC-CG--C-C-PC C
```

Fig. 1. Amino acid sequences of the products of plant genes with sequence similarity to metallothionein genes predicted from cDNA sequences, genomic DNA sequences or DNA sequences obtained from PCR mediated cloning. [a]Robinson and co-workers (unpublished data); [b]Evans et al. (1990); [c]de Miranda et al. (1990); [d]Kawashima et al. (1991); [e]de Framond (1991); [f]Okamura et al. (1991).

total RNA and poly(A^{-})RNA isolated from roots of 14 day old plants (Fig. 2A, B). Transcripts were less abundant in poly(A^{-})RNA from etiolated leaves and were barely detectable in isolates from leaves grown in the light. Weak hybridization to slightly smaller transcripts, which may be the product of a *PsMT* gene other than $PsMT_A$, was detected in developing cotyledons (Fig. 2B). *PsMT* transcripts were not detected in RNA isolated from embryonic pea radicles but increased in abundance in the roots during germination (Fig. 2C). ^{32}P-labelled $PsMT_A$ cDNA did not hybridize to sections of embryonic radicle using squash-blotting techniques (data not presented), but did hybridize to squash-blots of 14 day old pea roots (Fig. 3). These data were consistent with northern analyses (Fig. 2C). Hybridization was observed within the parenchymatous cortex, but not in the stele or the lateral root tips.

Observed increases in the abundance of $PsMT_A$ transcripts in pea roots during the first 14 days after germination could suggest that this gene is subject to developmental control. Alternatively, analogy to the expression of *ids1* in barley, might indicate that the pea plants used in these experiments were subject to increasing Fe deficiency during the course of root development, causing activation of the $PsMT_A$ gene. This explanation seems feasible since the stringent growth conditions used by Grusak et al. (1990) to avoid activation of Fe efficiency mechanisms in *P. sativum* were not employed in these

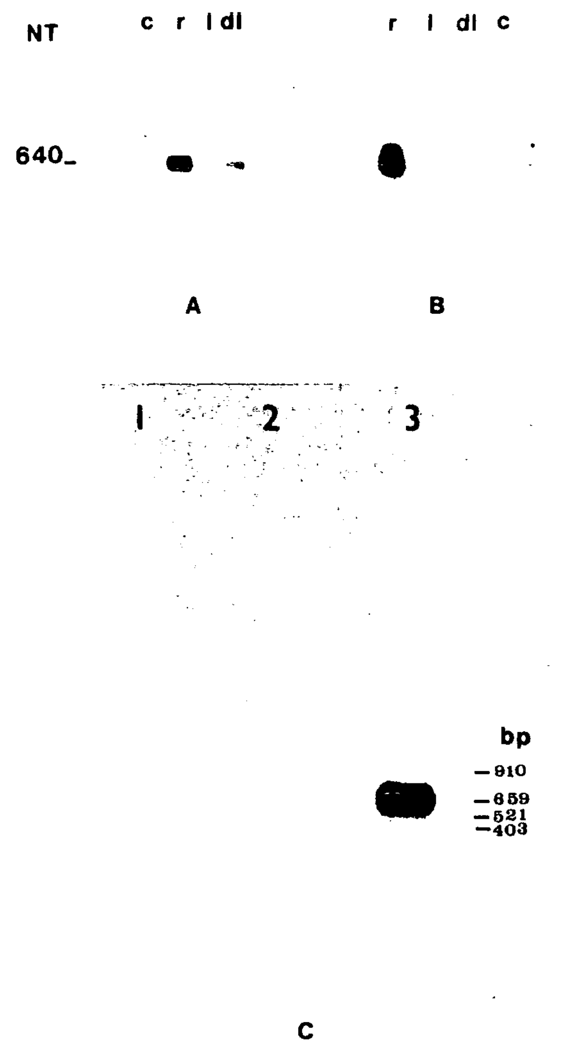


Fig. 2. Detection of *PsMT* mRNA in extracts from pea organs. Northern blot analysis was performed for samples of poly(A^{-})RNA (panel **A**) and total RNA (panel **B**) isolated from different organs. Agarose gels (1.5%) were loaded with total RNA (5 μg/lane), or poly(A^{-})RNA (2 μg/lane), from developing cotyledons 15 days after flowering (lane c), roots 14 days after germination (lane r), leaves (lane l), etiolated leaves (lane dl). Panel C shows the detection of *PsMT* mRNA in extracts from pea roots at different stages of development: radicle 13 days after flowering (lane 1), from roots 4 days after germination (lane 2) and roots 14 days after germination (lane 3).

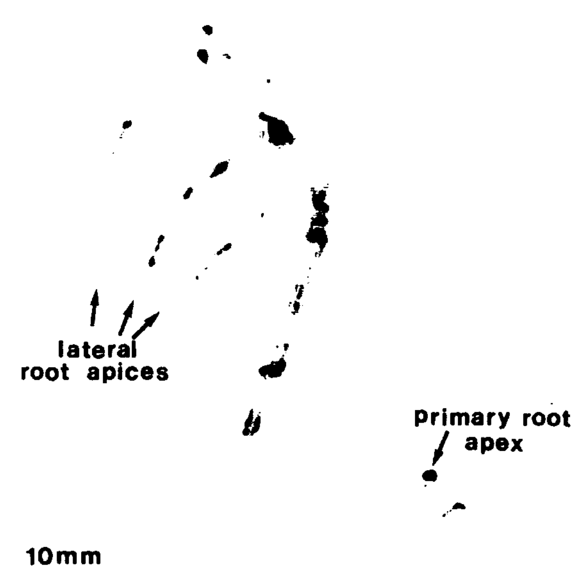


Fig. 3. Detection of *PsMT* mRNA on a tissue blot of pea roots 14 days after germination.

studies. It is also noted that preliminary results (data not shown) reveal that the Cu content of pea roots increased when plants were grown under conditions of Fe deficiency. It is possible that MT-like plant genes may be expressed under low Fe conditions in response to such elevated levels of Cu and function to detoxify excess of this metal ion. Clearly, further experiments are required to examine the pattern of expression of *PsMT* genes in pea roots following growth under different regimes of Fe availability.

In order to examine the metal binding characteristics of its product, the $PsMT_A$ gene has been expressed in *E. coli*. Table 1 shows the amount of individual trace metals associated with GST-$PsMT_A$ fusion protein, and with GST alone, following purification from lysates of 500 mL cultures grown in media supplemented with 500 μM Fe, Zn, Cu or Cd. A greater amount of metal associates with the fusion protein than with GST alone, suggesting that metal ions preferentially bind to the $PsMT_A$ portion of the protein. However, these ratios of metal to protein may not be reliable estimates of metal:protein stoichiometries, since differences could exist in the reactivities of GST, $PsMT_A$ and BSA with Coomassie blue, the reagent used to estimate protein concentrations. Furthermore, in subsequent experiments with replicated 50 mL cultures of *E. coli*, it was observed that isolates of GST-$PsMT_A$ from Cu exposed cells contained impurities which can also interfere with the estimated ratio of metal to protein (Tommey et al., 1991). The amount of Fe associated with the purified fusion protein was considerably less than the other metals. Explanations for this include the possibilities that 1) the Fe-protein binding may be extremely air sensitive, as observed for reconstituted animal Fe^{2+}-MT (Kägi and Kojima, 1987); 2) the intracellular levels of available Fe did not increase to allow enhanced Fe-binding; 3) the protein does not have a high affinity for Fe and preferentially binds other metal ions.

Table 1. Total amounts of metal and purified protein in the void fraction (3.5 mL) eluted from columns of Sephadex G-25 (PD-10, Pharmacia). Protein was purified by glutathione-affinity-chromatography from extracts of *E. coli* cells expressing either GST or GST-$PsMT_A$ fusion protein. Cells (500 mL) were grown in media supplemented with 500 μM of either Fe, Zn, Cu or Cd or without metal supplementation (figures in parentheses)

Purified protein		Metal (n moles)	Protein (n moles)	Ratio (metal:protein)
Fe	GST	16.49	167.7	0.1
		(0.0)	(111.9)	(0.0)
	GST-$PsMT_A$	19.75	71.2	0.28
		(0.0)	(72.1)	(0.0)
Zn	GST	28.28	76.4	0.37
		(1.35)	(111.9)	(0.01)
	GST-$PsMT_A$	397.8	66.4	5.99
		(55.98)	(72.1)	(0.78)
Cu	GST	92.6	146.4	0.63
		(18.07)	(111.9)	(0.16)
	GST-$PsMT_A$	414.1	66.0	6.27
		(54.22)	(72.1)	(0.75)
Cd	GST	37.6	144.0	0.26
		(7.58)	(111.9)	(0.07)
	GST-$PsMT_A$	299.9	42.4	7.07
		(11.36)	(72.1)	(0.16)

Many previous attempts to isolate MTs from plant roots have employed material excised from plants grown in the presence of highly elevated concentrations of trace metals. However, under these conditions plants accumulate phytochelatins (for reviews see Steffens, 1990; Rauser, 1990; Robinson, 1990), the presence of which may confound attempts to purify any MT protein which may also be present but possibly in much

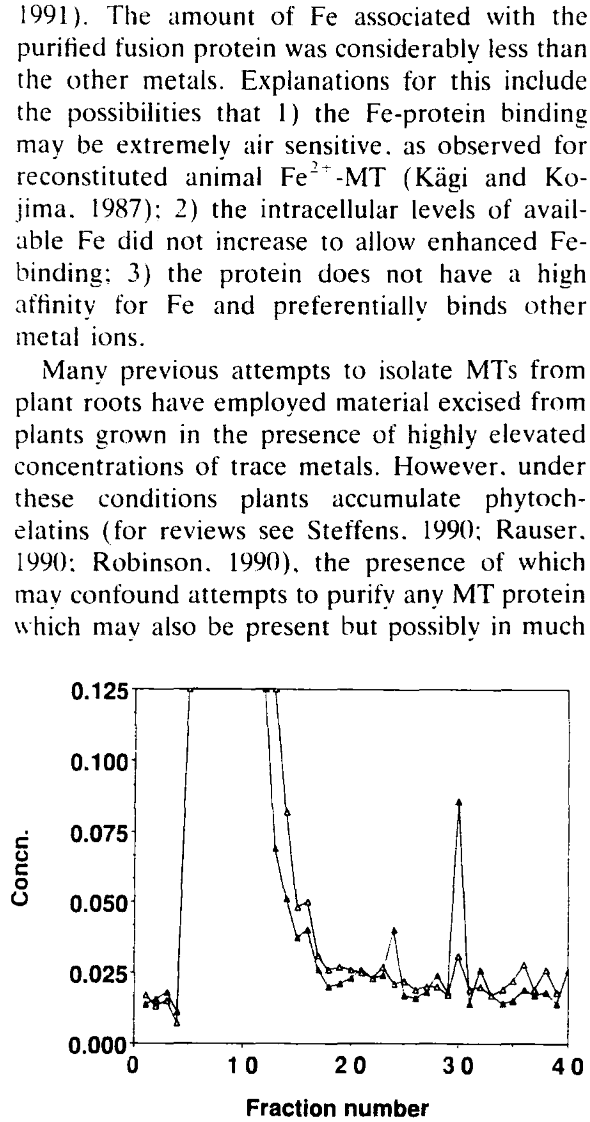


Fig. 4. Reversed phase HPLC analysis of phytochelatins in extracts from pea roots grown in the presence (▲) or absence of Cd (△). The thiol-content of fractions is shown (in μg glutathione equivalents mL^{-1}). The initial large peak (off-scale) is due to the presence of β-mercaptoethanol in the extraction buffer while the two lesser peaks (fractions 24 and 30) detected in extract from Cd-exposed material correspond to diagnostic retention times for phytochelatins.

lower quantities. Figure 4 shows the detection of phytochelatins in the roots of pea following growth in the presence of 50 μM Cd, consistent with previous observations that these polypeptides are present in pea. The assay was not sufficiently sensitive to clearly detect the lower levels of these polypeptides, which we assume to be present in pea roots grown in the absence of additional metal ions, although a small peak (putatively phytochelatin) is present in fraction 29 of the control extract (Fig. 4).

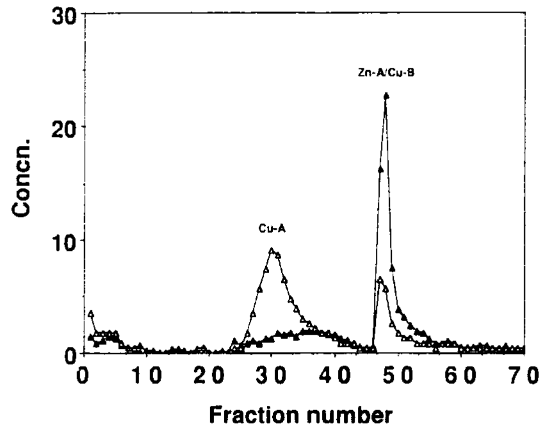


Fig. 5. Anion exchange chromatography on DEAE Sephadex of a 'constitutive' extract from pea roots. The Cu (△) and Zn (▲) content of fractions is shown in μM. The matrix was sequentially eluted with buffer containing 10 mM Tris (fractions 1 to 20), 100 mM Tris (fractions 21 to 40) and 300 mM Tris (fraction 41 onwards).

The data presented here show that $PsMT_A$ transcripts are abundant in roots grown without supra-optimal trace metal ion supplementation and indicate that the $PsMT_A$ protein will bind Cu and Zn following expression in *E. coli*. Purification of low M_r constitutive Cu and Zn complexes has therefore been initiated from pea roots with the eventual aim of establishing whether or not the ligands include the products of *PsMT* genes. Heat stable material from acetone precipitates of pea root extracts was crudely fractionated by stepwise elution from DEAE Sephadex (Fig. 5). Fractions eluted in buffer containing 100 mM Tris included a Cu complex which has been designated Cu-A, while fractions eluted in buffer containing 300 mM Tris included both Zn and Cu complexes, designated Zn-A and Cu-B, respectively. Figure 6 shows the subsequent frac-

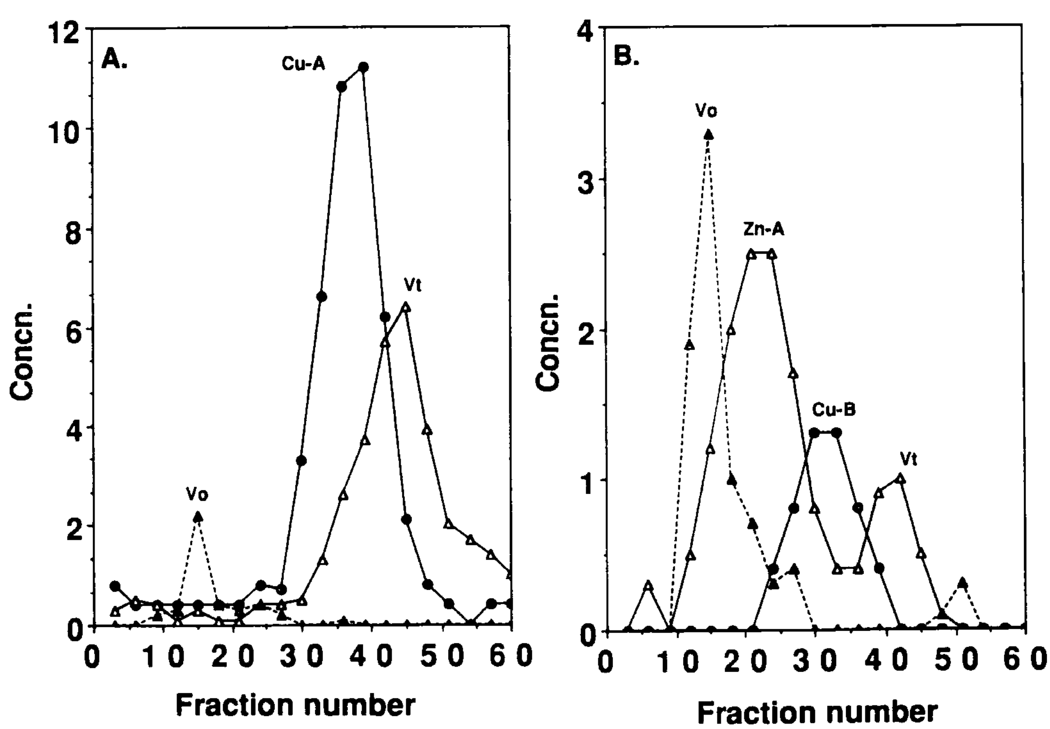


Fig. 6. Gel filtration on Sephadex G-50 of pooled fractions containing Cu-A (panel A) and Cu-B and Zn-A (panel **B**) after anion exchange chromatography. Every third fraction was analysed for protein (▲) in $\mu g\,mL^{-1}$, Zn (△) and Cu (●) in μM.

tionation of these complexes on Sephadex G-50. Cu-A and Cu-B have similar low M_r, while Zn-A is apparently larger than the Cu complexes. Antibodies are currently being raised to the expressed GST-PsMT$_A$ fusion protein. An initial aim will be to establish whether or not fractions containing Cu-A, Cu-B or Zn-A include antigens that cross-react with this antibody, indicative of metal-binding by the products of *PsMT* genes.

Acknowledgements

This work was supported by research grant PG12/519(PMB) from the Agricultural and Food Research Council. N. J. R. is a Royal Society University Research Fellow.

References

Delhaize E, Jackson P J, Lujan L D and Robinson N J 1989 Poly(γ-glutamylcysteinyl)glycine synthesis and aggregation with cadmium: Role in cadmium-tolerance. Plant Physiol. 89, 700–706.

Evans I M, Swinhoe R, Gatehouse L N, Gatehouse J A and Boulter D 1988 Distribution of root mRNA species in other vegetative organs of pea (*Pisum sativum* L.). Mol. Gen Genet. 214, 153–157.

Evans I M, Gatehouse L N, Gatehouse J A, Robinson N J and Croy R R D C 1990 A gene from pea (*Pisum sativum* L.) with homology to metallothionein genes. FEBS Lett. 262, 29–32.

Feinberg A P and Vogelstein B 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6–13.

Grill E, Winnaker E L and Zenk M H 1985 Phytochelatins: The principal heavy-metal complexing peptide of higher plants. Science 230, 574–576.

Grill E, Winnaker E-L, and Zenk M H 1987 Phytochelatins, a class of heavy-metal binding peptides from plants, are functionally analogous to metallothioneins. Proc. Natn. Acad. Sci. 84, 439–443.

Grusak M A, Welch R M and Kochian L V 1990 Physiological characterization of single-gene mutation of *Pisum sativum* exhibiting excess iron accumulation. Plant Physiol. 93, 976–981.

Hanley-Bowdoin L and Lane B G 1983 A novel protein programmed by the mRNA conserved in dry wheat embryos: The principal site of cysteine incorporation during early germination. Eur. J. Biochem 135, 9–15.

Jackson P J, Unkefer C J, Doolen J A, Watt K and Robinson N J 1987 Poly(γ-glutamylcysteinyl)glycine: Its role in cadmium resistance in plant cells. Proc. Natn. Acad. Sci. 84, 6619–6623.

Kägi J H R and Kojima Y 1987 Chemistry and Biochemistry of metallothionein. *In* Metallothionein II. Eds. J H R Kägi and Y Kojima. pp 63–80. Birkhäuser Verlag, Basel-Boston.

Kägi J H R and Schäffer A 1988 Biochemistry of metallothionein. Biochemistry 27, 8509–8515.

Kawashima I, Inokuchi Y, Chino M, Kimura M and Shimizu N 1991 Isolation of a gene for a metallothionein-like protein from soybean. Plant Cell Physiol. 32, 913–916.

Kondo N, Isobe M, Imai K and Goto T 1985 Synthesis of metallothionein-like peptides cadystin A and B occurring in fission yeast, and their isomers. Agric Biol. Chem. 49, 71–83.

Lane B, Kajioka R and Kennedy T 1987 The wheat germ E_c protein is a zinc-containing metallothionein Biochem. Cell Biol. 65, 1001–1005.

McMasters G K and Carmichael G G 1977 Analysis of single-stranded and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natn. Acad. Sci. 74, 4835–4838.

de Framond A J 1991 A metallothionein-like gene from maize (*Zea mays*) Cloning and characterization. FEBS Lett. 290, 103–106.

de Miranda J R, Thomas M A, Thurman D A and Tomsett A B 1990 Class I metallothionein genes from the flowering plant *Mimulus guttatus*. FEBS Lett. 260, 277–280.

Okumura N, Nishizawa N-K, Umehara Y and Mori S 1991 An iron deficiency-specific cDNA from barley roots having two homologous cysteine-rich MT domains. Plant Molec. Biol. 17, 531–533.

Palmiter R D 1987 Molecular biology of metallothionein gene expression. *In* Metallothionein II. Eds. J H R Kägi and Y Kojima. pp 63–80. Birkhäuser Verlag, Basel-Boston.

Pressler U, Staudenmaier H, Zimmermann L and Braun V (1988) Genetics of the iron dicitrate transport system of *Escherichia coli*. J. Bacteriol. 170 2716–2724.

Rauser W E 1990 Phytochelatins. Annu. Rev. Biochem. 59, 61–86.

Robinson 1990 Metal-binding polypeptides in plants. *In* Heavy-metal tolerance in plants. Ed. A J Shaw. pp 195–214. CRC Press, Boca Raton FL.

Robinson N J, Gupta A, Fordham-Skelton A P, Croy R R D, Whitton B A and Huckle J W 1990 Prokaryotic metallothionein gene characterisation and expression: Chromosome crawling by ligation-mediated PCR. Proc. R. Soc. Lond. B. 242, 241–247.

Smith D B and Johnson K S 1988 Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. Gene 67, 31–40.

Steffens J C 1990 The heavy metal-binding peptides of plants. A. Rev. Pl. Physiol. Pl. Mol. Biol. 41, 553–575.

Tommey A M, Shi J, Lindsay W P, Urwin P E and Robinson N J 1991 Expression of the pea gene *PsMT*$_A$ in *E. coli*: Metal-binding properties of the expressed protein. FEBS Lett. 292, 48–52.

Tsujikawa K, Imai T, Katutani M, Kayamori Y, Mimura T, Otaki N, Kimura M, Fukuyama R and Shimizu N 1991 Localization of metallothionein in nuclei of growing primary cultured adult rat hepatocytes. FEBS Lett. 283, 239–242.